

Run on: June 8, 2001, 06:21:22 ; Search time 217.33 Seconds  
(without alignments)  
53.723 Million cell updates/sec

Title:	US-09-507-242-3
perfect score:	20
Sequence:	1 ccagatggtataagtagaac 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1350532

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%

```
Database : N_Geneseq_0401:*
1. /STD56/acqdata/geneseq/geneseqn/NA1980.DAT:*
```

1. N\_Geneseq\_0401: \*
- 1: /SID56/gcgdata/geneseq/geneseqn/NA1980.DAT: \*
- 2: /SID56/gcgdata/geneseq/geneseqn/NA1981.DAT: \*
- 3: /SID56/gcgdata/geneseq/geneseqn/NA1982.DAT: \*
- 4: /SID56/gcgdata/geneseq/geneseqn/NA1983.DAT: \*
- 5: /SID56/gcgdata/geneseq/geneseqn/NA1984.DAT: \*
- 6: /SID56/gcgdata/geneseq/geneseqn/NA1985.DAT: \*
- 7: /SID56/gcgdata/geneseq/geneseqn/NA1986.DAT: \*
- 8: /SID56/gcgdata/geneseq/geneseqn/NA1987.DAT: \*
- 9: /SID56/gcgdata/geneseq/geneseqn/NA1989.DAT: \*
- 10: /SID56/gcgdata/geneseq/geneseqn/NA1990.DAT: \*
- 11: /SID56/gcgdata/geneseq/geneseqn/NA1991.DAT: \*
- 12: /SID56/gcgdata/geneseq/geneseqn/NA1992.DAT: \*
- 13: /SID56/gcgdata/geneseq/geneseqn/NA1993.DAT: \*
- 14: /SID56/gcgdata/geneseq/geneseqn/NA1994.DAT: \*
- 15: /SID56/gcgdata/geneseq/geneseqn/NA1995.DAT: \*
- 16: /SID56/gcgdata/geneseq/geneseqn/NA1996.DAT: \*
- 17: /SID56/gcgdata/geneseq/geneseqn/NA1997.DAT: \*
- 18: /SID56/gcgdata/geneseq/geneseqn/NA1998.DAT: \*
- 19: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT: \*
- 20: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT: \*
- 21: /SID56/gcgdata/geneseq/geneseqn/NA2001.DAT: \*
- 22: /SID56/gcgdata/geneseq/geneseqn/NA2001.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	20	100.0	20	21	A50293	Candida albicans C
2	100.0	989	21	21	A50292	Candida albicans C
3	16.4	82.0	19	21	A83104	cdk7 ribozyme bind
c	15.8	79.0	2224	18	V74796	Staphylococcus aur
4	15.4	77.0	19	21	A83103	cdk7 ribozyme bind
5	15.2	76.0	470	18	T75261	Nucleotide sequenc
c	15.2	76.0	981	21	C44949	Arabidopsis thalia
c	15.2	76.0	1434	21	C45997	Arabidopsis thalia
c	15.2	76.0	1606	13	Q28162	Human R3BP. Homo
c	15.2	76.0	1792	21	C46375	Arabidopsis thalia
c	15.2	76.0	1797	21	C34601	Arabidopsis thalia

C	12	15.2	76.0	189.2	21	C52084
C	13	15.2	76.0	229.9	21	T61274
C	14	15.2	76.0	4780	16	O94253
C	15	15	75.0	63	20	V87836
C	16	14.8	74.0	42	11	004070
C	17	14.8	74.0	42	14	042527
C	18	14.8	74.0	312	19	V30233
C	19	14.8	74.0	792	21	C54232
C	20	14.8	74.0	923	21	C35559
C	21	14.8	74.0	1563	21	A93072
C	22	14.8	74.0	2058	21	C46199
C	23	14.8	74.0	4247	16	O75303
C	24	14.8	74.0	4471	20	X00915
C	25	14.8	74.0	5136	21	Z98332
C	26	14.8	72.0	6828	16	O75302
C	27	14.4	72.0	1132	21	C36483
C	28	14.4	72.0	1626	21	C74280
C	29	14.4	72.0	2220	21	C79691
C	30	14.4	72.0	65921	21	T89046
C	31	14.4	72.0	580073	18	T58840
C	32	14.4	71.0	227	21	A44514
C	33	14.2	71.0	251	19	X11239
C	34	14.2	71.0	251	19	X11240
C	35	14.2	71.0	260	21	C43927
C	36	14.2	71.0	276	19	V26652
C	37	14.2	71.0	298	21	A01212
C	38	14.2	71.0	304	21	A43457
C	39	14.2	71.0	310	21	C81838
C	40	14.2	71.0	363	21	C81766
C	41	14.2	71.0	514	21	C37256
C	42	14.2	71.0	585	21	C52247
C	43	14.2	71.0	65158	21	X61588
C	44	14.2	71.0	639	22	C84300
C	45	14.2	71.0	651	20	X61557

## ALIGNMENTS

RESULT	1
A50293	
ID	A50293 standard; DNA; 20 BP.

20-NOV-2000 (first entry)

Candida albicans CaESS1 gene-specific primer OW-216

CaESS1: ESS1; infection; diagnosis; therapy; antifungal; fungicide  
KW antifungative; cytostatic; antitumour; PCR primer; ss.

*Candida albicans*

PN WO2000050561-A2.

PD 31-AUG-2000

AA  
PF 18-FEB-2000; 2000WO-US04203.

XX 23-FEB-1999; 99US-0121246  
PR

PA . (HEAL-) HEALTH RES INC.

xx  
PI Hanes SD, Devasahayam G, Chaturvedi V,

DR WPI; 2000-565453/52.

XX	Novel <i>Candida albicans</i> gene, Caess1 useful for identifying compounds
PI	that specifically bind to and/or inhibit Caess1 and thus for treating
PT	life-threatening fungal infections and other life-threatening fungal

PT Candida albicans infections -  
PT  
XX

PS Claim 6; Page 38; 51pp; English.

CC This is the nucleotide sequence of PCR primer OW-216, which is  
CC specific to the CaESS1 gene (see A50292), the Candida albicans  
CC homologue of Saccharomyces cerevisiae ESS1. OW-216 can be used  
CC with primer OW-221 (see A50294) for the specific amplification of  
CC CaESS1 DNA; they do not amplify ESS1 DNA or PIN1 (the human  
CC homologue of ESS1) DNA. Detection of CaESS1 nucleic acids can  
CC be used as a means for diagnosing C. albicans infection.

XX Sequence 20 BP; 8 A; 3 C; 5 G; 4 T; 0 other;

QY 1 ccagatggtataagtagaac 20  
Db 1 ccagatggtataagtagaac 20

Query Match 100.0%; Score 20; DB 21; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2

A50292  
ID A50292 standard; DNA; 989 BP.  
XX  
AC A50292;  
XX  
DT 20-NOV-2000 (first entry)  
XX  
DE Candida albicans CaESS1 gene.  
XX  
KW CaESS1; ESS1; Infection; diagnosis; therapy; antifungal; fungicide;  
KW antiproliferative; cytostatic; antitumour; ds.  
XX  
OS Candida albicans.  
XX  
FH Key Location/Qualifiers  
FT CDS 256..789  
FT /tag="a  
XX  
PN WO200050561-A2.  
XX  
PD 31-AUG-2000.  
XX  
PE 18-FEB-2000; 2000WO-US04203.  
XX  
PR 23-FEB-1999; 99US-0121246.  
XX  
PA (HEAL-) HEALTH RES INC.  
XX  
PI Hanes SD, Devasahayam G, Chaturvedi V;  
XX  
DR WPI: 2000-565453/52.  
DR P-PSDB; Y95876.  
XX  
PT Novel Candida albicans gene, CaESS1 useful for identifying compounds  
PT that specifically bind to and/or inhibit CaESS1 and thus for treating  
PT Candida albicans infections and other life-threatening fungal  
PT infections -  
XX  
XX Claim 2; Fig 1A; 51pp; English.

CC This is the complete nucleotide sequence of the CaESS1 gene from  
CC Candida albicans. The gene encodes a 177-amino acid protein,  
CC CaESS1 (see Y95876), which is the C. albicans homologue of  
CC Saccharomyces cerevisiae ESS1 (428 amino acid identity). The  
CC CaESS1 gene was isolated from a C. albicans genomic DNA library by  
CC functional complementation of a temperature-sensitive S. cerevisiae  
CC strain, ess1-194pts. CaESS1 nucleic acids, especially CaESS1-specific  
CC primers (see A50293-94) and probes to determine the presence of C.  
CC albicans in a sample or specimen. CaESS1 protein is a target for  
CC screening for antifungal and inhibitor compounds, useful for

CC treating or preventing C. albicans infections. CaESS1 DNA can also  
CC be used to generate diagnostic probes or primers for replicating or  
CC cloning C. albicans DNA. A CaESS1 inhibitor can be a compound  
CC which selectively inhibits growth of S. cerevisiae not containing an  
CC endogenous ESS1 gene but rather CaESS1 and uninduced PIN1 (a human  
CC homologue of ESS1) and/or preferably does not inhibit induced PIN1,  
CC e.g. does not inhibit S. cerevisiae not containing an endogenous  
CC ESS1 but rather induced PIN1. Compositions which inhibit PIN1  
CC are useful antiproliferatives e.g. antineoplastics, antitumour  
CC agents or anticancer agents.

XX Sequence 989 BP; 340 A; 174 C; 190 G; 285 T; 0 other;

QY 1 ccagatggtataagtagaac 20  
Db 520 ccagatggtataagtagaac 539

Query Match 100.0%; Score 20; DB 21; Length 989;  
Best Local Similarity 100.0%; Pred. No. 0.54;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3

A83104  
ID A83104 standard; DNA; 19 BP.  
XX  
AC A83104;  
XX  
DT 04-DEC-2000 (first entry)  
XX  
DE cdk7 ribozyme binding site #25.  
XX  
KW Ribozyme; hairpin; hammerhead; gene therapy; vasotropic;  
KW restenosis; ss.  
XX  
OS Mammalia.  
XX  
PN WO200032765-A2.  
XX  
PD 08-JUN-2000.  
XX  
PE 06-DEC-1999; 99WO-US28772.  
XX  
PR 04-DEC-1998; 98US-0110954.  
XX  
PA (IMMU-) IMMUSOL INC.  
XX  
PI Tritz R, Welch PJ, Barber JR, Robbins JM;  
XX  
DR WPI: 2000-412314/35.  
XX  
PT New hairpin and hammerhead ribozyme for inhibiting restenosis, cleaves  
PT RNA encoding a cyclin or cell-cycle dependent kinase other than CDK1,  
PT PCNA and Cyclin B1 -  
XX  
XX Disclosure; Page 56; 109pp; English.

CC The present invention relates to a hairpin or hammerhead ribozyme,  
CC designed to cleave RNA encoding a cyclin or cell-cycle dependent kinase  
CC other than cell-cycle dependent kinases CDK1, PCNA and Cyclin B1.  
CC Representative examples of ribozyme recognition sites are given in  
CC A82415 to A86787. The ribozyme of the invention is useful for  
CC inhibiting restenosis by introduction of the ribozyme into cells.  
CC The ribozyme is resistant to endonuclease activity and hence is  
CC efficient in restenosis treatment.

XX Sequence 19 BP; 9 A; 2 C; 4 G; 4 T; 0 other;

QY 1 ccagatggtataagtagaac 20  
Db 520 ccagatggtataagtagaac 539

Query Match 82.0%; Score 16.4; DB 21; Length 19;  
Best Local Similarity 94.4%; Pred. No. 24;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 3 agatggtataagtagaac 20  
 |||||  
 Db 1 agatggtataaataagAAC 18

## RESULT 4

V74796/c  
 ID V74796 standard; DNA: 2224 BP.

AC V74796;  
 XX

DT 16-MAR-1999 (first entry)  
 XX

DE Staphylococcus aureus contig SEQ ID #485.  
 XX

KM Computer readable medium: vaccine; S.aureus infection; immunodetection;  
 XX cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KM skin infection; surgical wound infection; scalded skin syndrome;  
 KM toxic shock syndrome; ds.

XX Staphylococcus aureus.  
 OS

XX Key Location/Qualifiers  
 FH misc-feature 1681..1740

FT /tag- a  
 FT /note- a

FT "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT EP786519-A2.  
 XX

PN 30-JUL-1997.  
 XX

PD 07-JAN-1997; 97EP-0100117.  
 XX

PF 05-JAN-1996; 96US-0009861.  
 XX

PR (HUMA-) HUMAN GENOME SCI INC.  
 XX

PA Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
 PI Rosen CA;  
 PI

XX WPI: 1997-374922/35.  
 DR

XX Polynucleotide(s) and proteins derived from Staphylococcus aureus  
 PT stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines

PS Claim 1; Page 1412-1413; 3271pp; English.  
 XX

XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC that S.aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against S.aureus infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.  
 XX

Sequence 2224 BP; 785 A; 291 C; 435 G; 648 T; 65 other;

Query Match 79.0%; Score 15.8; DB 18; Length 2224;  
 Best Local Similarity 89.5%; Pred. No. 70;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 cagatggtataagtagaac 20  
 |||||  
 Db 796 CAGATGCTTTAAGTACAC 778

## RESULT 5

A83103  
 ID A83103 standard; DNA: 19 BP.

AC A83103;  
 XX

DT 04-DEC-2000 (first entry)  
 XX

DE cdk7 ribozyme binding site #24.  
 XX

KM Ribozyme; hairpin; hammerhead; gene therapy; vasotropic;  
 KM restenosis; ss.

XX Mammalia.  
 OS

XX WO200032765-A2.  
 XX

PN 08-JUN-2000.  
 XX

PD 06-DEC-1999; 99WO-US28772.  
 XX

PF 04-DEC-1998; 98US-0110954.  
 XX

PR (IMMU-) IMMUSOL INC.  
 XX

PA Tritz R, Welch PJ, Barber JR, Robbins JM;  
 PI WPI: 2000-412314/35.  
 XX

DR New hairpin and hammerhead ribozyme for inhibiting restenosis, cleaves  
 PT RNA encoding a cyclin or cell-cycle dependent kinase other than CDK1,  
 PT PCNA and Cyclin B1  
 PT

XX Disclosure; Page 56; 109pp; English.  
 PS

XX The present invention relates to a hairpin or hammerhead ribozyme,  
 CC designed to cleave RNA encoding a cyclin or cell-cycle dependent kinase  
 CC other than cell-cycle dependent kinases CDK1, PCNA and Cyclin B1.  
 CC Representative examples of ribozyme recognition sites are given in  
 CC A82415 to A86787. The ribozyme of the invention is useful for  
 CC inhibiting restenosis by introduction of the ribozyme into cells.  
 CC The ribozyme is resistant to endonuclease activity and hence is  
 CC efficient in restenosis treatment.  
 CC

XX Sequence 19 BP; 11 A; 0 C; 4 G; 4 T; 0 other;  
 SQ

Query Match 77.0%; Score 15.4; DB 21; Length 19;  
 Best Local Similarity 94.1%; Pred. No. 76;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 3 agatggtataagtagaac 19  
 |||||  
 Db 3 agatggtataaataagAAC 19

## RESULT 6

T75261/c  
 ID T75261 standard; DNA: 470 BP.

AC T75261;  
 XX

DT 21-DEC-1998 (first entry)  
 XX

XX Nucleotide sequence encoding human RAD50 Intron 10.  
 DE ds; human; RAD50; DNA repair; tumour suppression; cancer; Septin-2;  
 XX central nervous system.  
 KM Homo sapiens.  
 XX MO92727284-A2.  
 XX 31-JUL-1997.  
 PD 24-JAN-1997; 97WO-US01299.  
 XX 17-JUL-1996; 96US-0687080.  
 PR 26-JAN-1996; 96US-0592126.  
 XX (GENE-) GENELABS TECHNOLOGIES INC.  
 PA Dolganov G;  
 XX WPI; 1997-393672/36.  
 DR Human tumour suppressor gene RAD50 - useful to detect  
 XX predisposition to, decrease risk of and treat cancer, also Septin-2  
 PT homologues  
 PS Claim 1; Page 111-112; 195pp; English.  
 XX The human RAD50 (hrRAD50) is involved in DNA repair and has tumour  
 CC suppression activity, can be used to detect predisposition to, decrease  
 CC the risk of or treat cancers, e.g. acute myeloid leukaemia,  
 CC myelodysplastic syndrome, therapy related myelodysplastic syndrome,  
 CC therapy related acute myeloid leukaemia, refractory anaemia or refractory  
 CC anaemia with excess blasts. Also disclosed in this invention is Human  
 CC Septin-2 homologues of which may be used as targets for cancer therapies  
 CC and central nervous system directed treatment methods, and to measure the  
 CC proliferative potential of selected cell types.  
 CC  
 SQ Sequence 470 BP; 154 A; 60 C; 60 G; 196 T; 0 other;

Query Match 76.0%; Score 15.2; DB 18; Length 470;  
 Best Local Similarity 85.0%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ccaagtggtataagtagaac 20  
 Db 37 CCAGCAGCATATATTAGAAC 18  
 ||||| ||||| |||||

RESULT 7  
 C44949/c  
 ID C44949 standard; DNA; 981 BP.  
 XX C44949;  
 AC  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 44735.  
 XX  
 KW Hybridisation assay; genetic mapping; gene expression control;  
 KW protein identification; signal transduction pathway;  
 KW metabolic pathway; promoter; termination sequence; ss.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX

PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126254.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 07-MAY-1999; 99US-0132487.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139464.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139859.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.



PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158332.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0162193.  
PR 29-OCT-1999; 99US-0162142.

Query Match 76.0%; Score 15.2; DB 21; Length 981;  
Best Local Similarity 85.0%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccagatggtataagtagaac 20  
||||| ||| ||||| |||||  
Db 650 CCAGAAGTTTAACTTGAAC 631

## RESULT 8

ID C45997 standard; DNA; 1434 BP.

AC C45997;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 48536.

XX Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PN EP103405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-01231825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140635.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.

```

PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match          76.0%; Score 15.2; DB 21; Length 1434;
Best Local Similarity 85.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 1 ccagatggtataagtagaac 20
    ||||| ||||| |||||
DB 1103 CCAGAGGCTTTAAGTTGAC 1084

```

```

RESULT 9
ID Q29162/c
AC Q29162 standard; DNA: 1606 BP.
XX Q29162;
DT 05-MAR-1993 (first entry)
XX
DE Human R3BP.
XX
KW Gastrin releasing peptide; GRP; bombesin; neuromedin B; NMB;
KW ranatensin; RNP; bombesin-like peptide; R1BP; R2BP; R3BP;
KW receptor; agonist; antagonist; ligand; antibody; cancer; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 172..1371
FT /tag= a
XX
PN WO9216623-A.
XX
PD 01-OCT-1992.
XX
PF 13-MAR-1992; 92WO-US02091.
XX
PR 15-MAR-1991; 91US-0670603.
PR 03-OCT-1991; 91US-0771332.
XX
PA (BERL.) BERLEX LAB INC.
PA (USSH.) US DEPT HEALTH & HUMAN SERVICE.
XX
PI Battey JF, Corlay MH, Fathi Z, Feldman RI, Harkin RN;
PI Slatery TK, Wada E, Wu JM;
XX
DR WPI: 1992-349208/42.

```

```

DR P-PSDB: R27511.
XX
XX Receptors for bombesin-like peptide(s) and their DNA - useful for
PT screening for agonists and antagonists of the receptor ligands,
PT also for treating or diagnosing cancer
XX
XX Disclosure: Page 147-149; 173pp; English.
XX
XX The DNA sequences encoding mouse R1BP, human R1BP, rat R2BP, human
CC R2BP and human R3BP are given in Q29158-62 respectively. The
CC receptor gene and encoded polypeptide are used for screening for
CC agonists and antagonists of the receptor ligands, for producing
CC diagnostic or therapeutic reagents, and for producing antibodies.
CC Hosts suffering from abnormal receptor function, e.g. proliferative
CC cell conditions such as cancers, may be treated.
CC The mouse GRP receptor was isolated from Swiss 3T3 fibroblasts and
CC sequenced. The sequence was used to design oligonucleotide probes
CC to isolate DNA encoding mouse GRP receptor from a Swiss 3T3 CDNA
CC library. This DNA was then used as a probe to isolate rat NMB
CC receptor, human GRP receptor, human NMB receptor and human R3BP
CC (incompletely characterised homologous putative receptor) from
CC DNA libraries.
XX
SQ Sequence 1606 BP; 428 A; 367 C; 323 G; 488 T; 0 other;

```

```

Query Match          76.0%; Score 15.2; DB 13; Length 1606;
Best Local Similarity 85.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 1 ccagatggtataagtagaac 20
    ||||| ||||| |||||
DB 87 CGAGATGCTATTGAGAGAC 68

```

```

RESULT 10
ID C46375/c
AC C46375 standard; DNA: 1792 BP.
XX C46375;
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 49913.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.

```

PR 30-APR-1999; 9905-0132407.  
PR 04-MAY-1999; 9905-0132484.  
PR 05-MAY-1999; 9905-0132485.  
PR 06-MAY-1999; 9905-0132486.  
PR 06-MAY-1999; 9905-0132487.  
PR 07-MAY-1999; 9905-0132483.  
PR 11-MAY-1999; 9905-0134256.  
PR 14-MAY-1999; 9905-0134219.  
PR 14-MAY-1999; 9905-0134221.  
PR 14-MAY-1999; 9905-0134370.  
PR 18-MAY-1999; 9905-0134768.  
PR 19-MAY-1999; 9905-0134941.  
PR 20-MAY-1999; 9905-0135124.  
PR 21-MAY-1999; 9905-0135353.  
PR 24-MAY-1999; 9905-0135629.  
PR 25-MAY-1999; 9905-0136021.  
PR 27-MAY-1999; 9905-0136392.  
PR 28-MAY-1999; 9905-0136782.  
PR 01-JUN-1999; 9905-0137222.  
PR 03-JUN-1999; 9905-0137528.  
PR 04-JUN-1999; 9905-0137502.  
PR 07-JUN-1999; 9905-0137724.  
PR 08-JUN-1999; 9905-0138094.  
PR 10-JUN-1999; 9905-0138540.  
PR 10-JUN-1999; 9905-0138847.  
PR 14-JUN-1999; 9905-0139119.  
PR 16-JUN-1999; 9905-0139452.  
PR 16-JUN-1999; 9905-0139453.  
PR 17-JUN-1999; 9905-0139492.  
PR 18-JUN-1999; 9905-0139454.  
PR 18-JUN-1999; 9905-0139455.  
PR 18-JUN-1999; 9905-0139456.  
PR 18-JUN-1999; 9905-0139457.  
PR 18-JUN-1999; 9905-0139458.  
PR 18-JUN-1999; 9905-0139459.  
PR 18-JUN-1999; 9905-0139460.  
PR 18-JUN-1999; 9905-0139461.  
PR 18-JUN-1999; 9905-0139462.  
PR 18-JUN-1999; 9905-0139463.  
PR 18-JUN-1999; 9905-0139750.  
PR 18-JUN-1999; 9905-0139763.  
PR 21-JUN-1999; 9905-0139817.  
PR 22-JUN-1999; 9905-0139899.  
PR 23-JUN-1999; 9905-0140353.  
PR 23-JUN-1999; 9905-0140354.  
PR 24-JUN-1999; 9905-0140695.  
PR 28-JUN-1999; 9905-0140823.  
PR 29-JUN-1999; 9905-0140991.  
PR 30-JUN-1999; 9905-0141287.  
PR 01-JUL-1999; 9905-0141842.  
PR 01-JUL-1999; 9905-0142154.  
PR 02-JUL-1999; 9905-0142055.  
PR 06-JUL-1999; 9905-0142390.  
PR 08-JUL-1999; 9905-0142803.  
PR 09-JUL-1999; 9905-0142977.  
PR 12-JUL-1999; 9905-0143542.  
PR 13-JUL-1999; 9905-0143562.  
PR 14-JUL-1999; 9905-0143624.  
PR 15-JUL-1999; 9905-0144005.  
PR 16-JUL-1999; 9905-0144085.  
PR 16-JUL-1999; 9905-0144086.  
PR 19-JUL-1999; 9905-0144325.  
PR 19-JUL-1999; 9905-0144331.  
PR 19-JUL-1999; 9905-0144332.  
PR 19-JUL-1999; 9905-0144333.  
PR 19-JUL-1999; 9905-0144334.  
PR 19-JUL-1999; 9905-0144335.  
PR 20-JUL-1999; 9905-0144352.  
PR 20-JUL-1999; 9905-0144632.  
PR 20-JUL-1999; 9905-0144884.  
PR 21-JUL-1999; 9905-0144814.  
PR 21-JUL-1999; 9905-0145086.

PR 21-JUL-1999; 9905-0145088.  
PR 22-JUL-1999; 9905-0145085.  
PR 22-JUL-1999; 9905-0145087.  
PR 22-JUL-1999; 9905-0145089.  
PR 22-JUL-1999; 9905-0145192.  
PR 23-JUL-1999; 9905-0145145.  
PR 23-JUL-1999; 9905-0145218.  
PR 26-JUL-1999; 9905-0145276.  
PR 27-JUL-1999; 9905-0145913.  
PR 27-JUL-1999; 9905-0145918.  
PR 27-JUL-1999; 9905-0145919.  
PR 28-JUL-1999; 9905-0145951.  
PR 02-AUG-1999; 9905-0146386.  
PR 02-AUG-1999; 9905-0146388.  
PR 02-AUG-1999; 9905-0146389.  
PR 03-AUG-1999; 9905-0147038.  
PR 04-AUG-1999; 9905-0147204.  
PR 04-AUG-1999; 9905-0147302.  
PR 05-AUG-1999; 9905-0147192.  
PR 05-AUG-1999; 9905-0147260.  
PR 06-AUG-1999; 9905-0147303.  
PR 06-AUG-1999; 9905-0147416.  
PR 09-AUG-1999; 9905-0147493.  
PR 09-AUG-1999; 9905-0147935.  
PR 10-AUG-1999; 9905-0148171.  
PR 11-AUG-1999; 9905-0148319.  
PR 12-AUG-1999; 9905-0148341.  
PR 13-AUG-1999; 9905-0148565.  
PR 13-AUG-1999; 9905-0148684.  
PR 16-AUG-1999; 9905-0149368.  
PR 17-AUG-1999; 9905-0149175.  
PR 18-AUG-1999; 9905-0149426.  
PR 20-AUG-1999; 9905-0149722.  
PR 20-AUG-1999; 9905-0149723.  
PR 20-AUG-1999; 9905-0149929.  
PR 23-AUG-1999; 9905-0149902.  
PR 25-AUG-1999; 9905-0150566.  
PR 26-AUG-1999; 9905-0150884.  
PR 27-AUG-1999; 9905-0151065.  
PR 27-AUG-1999; 9905-0151066.  
PR 27-AUG-1999; 9905-0151080.  
PR 27-AUG-1999; 9905-0151303.  
PR 30-AUG-1999; 9905-0151303.  
PR 31-AUG-1999; 9905-0151438.  
PR 01-SEP-1999; 9905-0151930.  
PR 07-SEP-1999; 9905-0152363.  
PR 10-SEP-1999; 9905-0153070.  
PR 13-SEP-1999; 9905-0153758.  
PR 15-SEP-1999; 9905-0154018.  
PR 16-SEP-1999; 9905-0154039.  
PR 20-SEP-1999; 9905-0154779.  
PR 22-SEP-1999; 9905-0155139.  
PR 23-SEP-1999; 9905-0155486.  
PR 24-SEP-1999; 9905-0155659.  
PR 24-SEP-1999; 9905-0156458.  
PR 29-SEP-1999; 9905-0156596.  
PR 04-OCT-1999; 9905-0157117.  
PR 05-OCT-1999; 9905-0157753.  
PR 06-OCT-1999; 9905-0157865.  
PR 07-OCT-1999; 9905-0158029.  
PR 08-OCT-1999; 9905-0158232.  
PR 12-OCT-1999; 9905-0158369.  
PR 13-OCT-1999; 9905-0159293.  
PR 13-OCT-1999; 9905-0159294.  
PR 14-OCT-1999; 9905-0159295.  
PR 14-OCT-1999; 9905-0159330.  
PR 14-OCT-1999; 9905-0159331.  
PR 14-OCT-1999; 9905-0159637.  
PR 14-OCT-1999; 9905-0159638.  
PR 18-OCT-1999; 9905-0159584.  
PR 21-OCT-1999; 9905-0160741.

PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 76.0%; Score 15.2; DB 21; Length 1792;  
Best Local Similarity 85.0%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ccagatgataagtagaac 20  
DB 1063 CCTGATGCTAAGTAGAAC 1044

## RESULT 11

C34601/c  
ID C34601 standard; DNA; 1797 BP.

XX C34601;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 7222.

XX Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;:

KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134376.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145213.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147304.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 23-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 13-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159684.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.

PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 25-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 26-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 76.0%; Score 15.2; DB 21; Length 1797;  
 Best Local Similarity 85.0%; Pred. No. 1.4e+02;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccagatgtataagtagaac 20  
 DB 1065 CCTGATGTCAATAGTAGAAC 1046

RESULT 12  
 ID C52084 standard; DNA; 1892 BP.  
 AC C52084;  
 XX  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 70346.  
 XX  
 KW Hybridisation assay; genetic mapping; gene expression control;  
 KW protein identification; signal transduction pathway;  
 KW metabolic pathway; promoter; termination sequence; ss.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140699.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 02-AUG-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0153363.  
PR 10-SEP-1999; 99US-0153707.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 26-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161361.

PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 76.0%; Score 15.2; DB 21; Length 1892;  
 Best Local Similarity 85.0%; Pred. No. 1.4e+02;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ccagatgataaagaac 20  
 ||||| ||| ||||| |||||  
 Db 1051 CCAGAGCTTTAAAGTTGAC 1032

## RESULT 13

T61274  
 ID T61274 standard; DNA; 2329 BP.

AC T61274;

DT 23-APR-1997 (first entry)

DE Penicillium decumbens epoxidation enzyme genomic DNA.

KW Epoxidase; epoxidation; antibiotic; phosphomycin; precursor;  
 (2)-1-propenyl phosphonic acid; ds.

OS Penicillium decumbens.

PH Key Location/Qualifiers  
 FT exon 1..1294  
 FT /\*tag- a  
 FT /number= 1  
 FT /codon.start= 1007  
 FT 1295..1421

FT Intron  
 FT /\*tag- b  
 FT /number= 1  
 FT 1422..1854

FT Exon  
 FT /\*tag- c  
 FT /number= 2  
 FT 1855..1916

FT Intron  
 FT /\*tag- d  
 FT /number= 2  
 FT 1917..2329

FT Exon  
 FT /\*tag- e

FT exon

FT exon

FT exon

FT exon

FT exon

FT exon

FT exon

FT exon

FT exon

FT exon

FT exon

FT exon

FT exon

FT exon

FT exon

Sequence 2329 BP; 597 A; 587 C; 623 G; 522 T; 0 other;

Query Match 76.0%; Score 15.2; DB 18; Length 2329;  
 Best Local Similarity 85.0%; Pred. No. 1.4e+02;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ccagatgataaagaac 20  
 ||||| ||| ||||| |||||  
 Db 1900 ctatagtgataaagaac 1919

## RESULT 14

O94253  
 ID O94253 standard; DNA; 4780 BP.

AC O94253;

DT 10-MAY-1996 (first entry)

DE Neuronal nitrogen monoxide synthase DNA.

KW nitrogen monoxide synthase; neuronal; treatment; prevention;  
 vascular disease; restenosis; ds.

OS Homo sapiens.

PH Key Location/Qualifiers  
 FT CDS 431..4732  
 FT /\*tag- a  
 FT /product= bNOS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

Sequence 4780 BP; 1159 A; 1413 C; 1292 G; 916 T; 0 other;

Query Match 76.0%; Score 15.2; DB 16; Length 4780;  
 Best Local Similarity 85.0%; Pred. No. 1.5e+02;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ccagatgataaagaac 20  
 ||||| ||| ||||| |||||  
 Db 3616 ccagatgataaagaac 3635



RESULT 15  
 ID V87826  
 AC V87826; standard; CDNA; 653 BP.  
 DT 12-FEB-1999 (first entry)  
 DE EST clone EW13.  
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;  
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
 KW chemokinesis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
 OS Homo sapiens.  
 PN W03845437-A2.  
 PD 15-OCT-1998.  
 PF 10-APR-1998; 98WO-US06956.  
 PR 10-APR-1997; 97US-0837312.  
 PA (GEMV ) GENETICS INST INC.  
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
 PI Racie LA, Spaulding V, Treacy M;  
 DR WPI: 1999-070078/06.  
 PT New polynucleotides encoding human secreted proteins - derived from  
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
 PT ovary, pituitary, retina and colon CDNA libraries  
 PS Claim 1; Page 192-193; 641pp; English.  
 XX The present sequence represents an expressed sequence tag (EST), and is  
 CC a polynucleotide of the invention. The polynucleotides of the invention  
 CC are all secreted EST sequences isolated from a variety of human tissue  
 CC sources. The EST sequences and proteins encoded by them are predicted to  
 CC have useful biological activities which would make them suitable for  
 CC treating, preventing or ameliorating medical conditions in humans and  
 CC animals, although no supporting data is given. Suggested activities  
 CC include nutritional activity, immune stimulating or suppressing activity,  
 CC haematopoiesis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC activity. The EST sequences are also stated to be useful for gene  
 CC therapy.  
 SQ Sequence 653 BP; 210 A; 108 C; 147 G; 188 T; 0 other;

Query Match 75.0%; Score 15; DB 20; Length 653;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 tggataagtagaac 20  
 ||||||||||||  
 DB 557 tggataagtagaac 571

Search completed: June 8, 2001, 06:21:25  
 Job time: 6149 sec

.....

**This Page Blank (uspto)**



**This Page Blank (uspto)**

provided by Dr. Bertrand Jordan. Library constructed and  
normalized by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 118 a 82 c 81 g 69 t

Query Match 79.0%; Score 15.8; DB 3; Length 350;  
Best Local Similarity 89.5%; Pred. No. 3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2 aatgacgggaacgttcg 20  
Db 234 AGTACCGGAACGTTCCG 252

## RESULT 13

LOCUS C69306 378 bp mRNA EST 23-SEP-1997  
DEFINITION C69306 Yuj1 Kohara unpublished cDNA Caenorhabditis elegans CDNA  
clone YK361c9 5', mRNA sequence.  
ACCESSION C69306  
VERSION C69306.1 GI:2430662  
KEYWORDS EST  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae  
; Rhabditidae; Pelodidae; Caenorhabditis.  
Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano  
M., Miyata, A. and Nishigaki, A.  
Expression map of the C. elegans genome  
Unpublished (1996)  
Contact: Yuj1 Kohara  
Gene Library Lab  
National Institute of Genetics  
Yata 1111, Mishima, Shizuoka 411, Japan  
Tel: 81-559-81-6854  
Fax: 81-559-81-6855  
Email: ykohara@lab.nig.ac.jp.  
Location/Qualifiers  
1. 378

FEATURES  
source  
/organism="Caenorhabditis elegans"  
/strain="CB1489 hlm-8(e1489)"  
/db\_xref="taxon:6239"  
/clone\_lib="YK361c9"  
/clone\_lib="Yuj1 Kohara unpublished CDNA"  
/sex="hermaphrodite, male"  
/tissue\_type="whole animal"  
/dev\_stage="varied"

BASE COUNT 103 a 94 c 75 g 106 t

Query Match 79.0%; Score 15.8; DB 175; Length 378;  
Best Local Similarity 89.5%; Pred. No. 3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2 aatgacgggaacgttcg 20  
Db 75 AATCAGGGAACGTTCCG 93

## RESULT 14

LOCUS A1026675 420 bp mRNA EST 07-JUL-1998  
DEFINITION A1026675 x1 Soares, testis\_NHR Homo sapiens CDNA clone IMAGE:145646  
3', mRNA sequence.  
ACCESSION A1026675  
VERSION A1026675.1 GI:3246163  
KEYWORDS EST  
SOURCE Homo sapiens.  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 420)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo  
, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
[www-bio.lnl.gov/bdrrp/image/image.html](http://www-bio.lnl.gov/bdrrp/image/image.html)  
Insert Length: 529 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 399.  
Location/Qualifiers  
1. 420

FEATURES  
source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:145646"  
/clone\_lib="Soares-testis\_NHR"  
/sex="male"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand CDNA  
was prepared from mRNA obtained from clonech laboratories  
, Inc., and primed with a Not I - oligo(dT) primer [5',  
TCGTACCATCTGCACTGGAGGCGCCCAATTTTITTTTITTTT 3'].  
Double-stranded CDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization to Cot5, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 113 a 103 c 105 g 99 t

Query Match 79.0%; Score 15.8; DB 15; Length 420;  
Best Local Similarity 89.5%; Pred. No. 3.1e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2 aatgacgggaacgttcg 20  
Db 266 AATGATGGAACGTTCCG 284

## RESULT 15

LOCUS AL381271 502 bp mRNA EST 03-AUG-2000  
DEFINITION M1BC019A02P1 M1BC Medicago truncatula CDNA clone M1BC019A0 T3, mRNA  
sequence.  
ACCESSION AL381271  
VERSION AL381271.1 GI:9681022  
KEYWORDS EST  
SOURCE barrel medic.  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
Fabales; Fabaceae; Papilionoideae; Medicago.

1 (bases 1 to 502)  
Journet, E.P., Crespeau, H., Van-Tuinen, D., Guzy, J., Jallion, O.,  
Miebel, A., Carreau, V., Chataigner, O., Kahn, D., Glanina, Zl-Pearson  
, V. and Gamas, P.

Medicago truncatula ESTs from endomycorrhizal roots  
Unpublished (2000)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: [sequef@genoscope.cns.fr](mailto:sequef@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)

**This Page Blank (uspto)**

m. misc.1 misc.fea. featu.

Query Match 84.0%; Score 16.8; DB 87; Length 220922;  
Best Local Similarity 90.0%; Pred. No. 1.9e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ccagatgataaagtagaac 20  
DB 127821 CCAGATGATTAATAAGAAC 127802

RESULT 11  
RNCIDKAK 989 bp mRNA ROD 04-JAN-1995  
LOCUS R.norvegicus mRNA for Cdk-activating kinase.  
DEFINITION X83579  
ACCESSION X83579.1 GI:619508  
VERSION X83579.1 GI:619508  
KEYWORDS Cdk-activating kinase.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 989)  
AUTHORS Wu, L. and Hall, F.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 989)  
AUTHORS Hall, F.  
TITLE Direct Submission  
COMMENT Submitted (19-DEC-1994) F. Hall, Children's Hospital Los Angeles,  
Division of Orthopaedic Surgery, 4650 Sunset Boulevard, Los  
Angeles, CA 90027, USA  
LOCATION/Qualifiers

FEATURES  
source 1. .989  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/tissue-type="testis"  
/cell-type="mixed"  
/cell\_line="rat testis CDNA"  
1. .941  
/gene="CAK1"  
/codon\_start=1  
/product="Cdk-activating kinase Cdk7"  
/protein\_id="CA58562.1"  
/db\_xref="GI:619509"  
/db\_xref="SWISS-PROT:P51952"  
/translation="ANNREKIDPLGSGQFATYKARDKNTNOIVAIRKIKLGRSEAK  
DGINRTALREIKLQELSHPTIGLDARFCHKNSISLVDPMDLEVLITKONSIVLT  
PSHIAVMIMTLOGLLEHLMHRLKNNLLDENGVKLADFGKLSGSPNNA  
YTHQVTVMYRARPPELLFGARMYGVGDMMAVGCIILAEILLRVPFLPGSDLDLQIRIT  
ETLGRTEEDQMDKSLPDYVTKSPFGIPLQIIFIAAGDDLELLQGLFLNPCTRI  
TASQLATKRYFSNRPGPQGLPRPCVPEALKEQ"

gene 1. .941  
CDS <1. .941  
/gene="CAK1"  
/codon\_start=1  
/product="Cdk-activating kinase Cdk7"  
/protein\_id="CA58562.1"  
/db\_xref="GI:619509"  
/db\_xref="SWISS-PROT:P51952"  
/translation="ANNREKIDPLGSGQFATYKARDKNTNOIVAIRKIKLGRSEAK  
DGINRTALREIKLQELSHPTIGLDARFCHKNSISLVDPMDLEVLITKONSIVLT  
PSHIAVMIMTLOGLLEHLMHRLKNNLLDENGVKLADFGKLSGSPNNA  
YTHQVTVMYRARPPELLFGARMYGVGDMMAVGCIILAEILLRVPFLPGSDLDLQIRIT  
ETLGRTEEDQMDKSLPDYVTKSPFGIPLQIIFIAAGDDLELLQGLFLNPCTRI  
TASQLATKRYFSNRPGPQGLPRPCVPEALKEQ"

BASE COUNT 291 a 217 c 235 g 246 t  
ORIGIN

Query Match 82.0%; Score 16.4; DB 94; Length 989;  
Best Local Similarity 94.4%; Pred. No. 3.8e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 3 agatgataaagtagaac 20  
DB 134 AGATGATTAATAAGAAC 151

KEYWORDS Cdk-activating kinase; protein kinase.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 1269)  
AUTHORS Wu, L., Yee, A., Liu, L., Carbonaro-Hall, D., Venkatesan, N., Tolo, V. T.  
and Hall, F. L.  
TITLE Molecular cloning of the human CAK1 gene encoding a  
cyclin-dependent kinase-activating kinase  
JOURNAL Oncogene 9 (7), 2089-2096 (1994)  
MEDLINE 94268852  
REFERENCE 2 (bases 1 to 1269)  
AUTHORS Hall, F. L.  
TITLE Direct Submission  
COMMENT Submitted (14-JAN-1994) F. L. Hall, Childrens Hospital Los Angeles,  
Division of Orthopaedic Surgery, 4650 Sunset Boulevard, Los  
Angeles, California 90027, USA  
REMARK Revised by [3]  
AUTHORS Wu, L.  
JOURNAL Direct Submission  
COMMENT Submitted (08-AUG-1996) L. Wu, Childrens Hospital Los Angeles,  
Division of Orthopaedic Surgery, 4650 Sunset Boulevard, Los  
Angeles, California 90027, USA  
On Aug 9, 1996 this sequence version replaced gi:509772.  
LOCATION/Qualifiers

FEATURES  
source 1. .1269  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue-type="connective"  
/cell-type="fibroblast"  
/cell\_line="M138 human fibroblast (ATCC)"  
/clone\_id="lambda CDNA"  
/clone="pBluescript Phagemid"  
44. .1084  
/gene="CAK1"  
/codon\_start=1  
/product="Cdk-activating kinase"  
/protein\_id="CA54508.1"  
/db\_xref="GI:1486357"  
/db\_xref="SWISS-PROT:P50613"  
/translation="VALDYKSRKRYREKIDPLGSGQFATYKARDKNTNOIVAIRKIK  
LGRSEAKDGINRTALREIKLQELSHPTIGLDARFCHKNSISLVDPMDLEVLIT  
KONSIVLTPSHIAVMIMTLOGLLEHLMHRLKNNLLDENGVKLADFGKLSGSD  
SEGSFNATYTHQVTVMYRARPPELLFGARMYGVGDMMAVGCIILAEILLRVPFLPGSD  
LDLQIRITETLGRTEEDQMDKSLPDYVTKSPFGIPLQIIFIAAGDDLELLQGLFLNPCTRI  
TASQLATKRYFSNRPGPQGLPRPCVPEALKEQ"

gene 44. .1084  
CDS 44. .1084  
/gene="CAK1"  
/codon\_start=1  
/product="Cdk-activating kinase"  
/protein\_id="CA54508.1"  
/db\_xref="GI:1486357"  
/db\_xref="SWISS-PROT:P50613"  
/translation="VALDYKSRKRYREKIDPLGSGQFATYKARDKNTNOIVAIRKIK  
LGRSEAKDGINRTALREIKLQELSHPTIGLDARFCHKNSISLVDPMDLEVLIT  
KONSIVLTPSHIAVMIMTLOGLLEHLMHRLKNNLLDENGVKLADFGKLSGSD  
SEGSFNATYTHQVTVMYRARPPELLFGARMYGVGDMMAVGCIILAEILLRVPFLPGSD  
LDLQIRITETLGRTEEDQMDKSLPDYVTKSPFGIPLQIIFIAAGDDLELLQGLFLNPCTRI  
TASQLATKRYFSNRPGPQGLPRPCVPEALKEQ"

BASE COUNT 425 a 227 c 277 g 340 t  
ORIGIN

Query Match 82.0%; Score 16.4; DB 92; Length 1269;  
Best Local Similarity 94.4%; Pred. No. 3.7e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 3 agatgataaagtagaac 20  
DB 199 AGATGATTAATAAGAAC 216

RESULT 13  
HSSTFMO15 1274 bp mRNA PRI 27-MAY-1997  
LOCUS H.sapiens mRNA for serine/threonine protein kinase MO15.  
DEFINITION Y13120  
ACCESSION Y13120.1 GI:2125815  
VERSION Y13120.1 GI:2125815  
KEYWORDS serine/threonine protein kinase.  
SOURCE human.  
ORGANISM Homo sapiens

**This Page Blank (uspto)**



FEATURES  
SOURCE

gene  
CDS

```

1. .1274
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="lung and thymus"
/clone="K2"
104. .1144
/gene="MO15"
104. .1144
/gene="MO15"

```

BASE COUNT  
ORIGIN

403 a 235 c 297 g 339 t

Query Match	82.0%	Score 16.4;	DB 93;	Length 1274;
Best Local Similarity	94.4%	Pred. No. 3.7e+02;		
Matches 17; Conservative		0; Mismatches 1;	Indels 0;	Gaps 0

```

OY      3 agatggtataagtagaac 20
          |||||
Db      259 AGATGCTATAAATAGAAC 276

```

RESULT 14  
AV076509

AK026509	LOCUS	DEFINITION
AK026509	1289 bp mRNA	PR1
Homo sapiens CDNA: FLJ22856 fls, clone KAT01479, highly similar to		
HSCA/CDC Homo sapiens Cdk mRNA for Cdk-activating kinase.		

AK026509 1289 bp mRNA P1  
Homo sapiens CDNA: FLJ22856 fls, clone KAT01479, highly similar to  
HSCAKCDK Homo sapiens CAK mRNA for CDK-activating kinase.

ACCESSION	AK026509
VERSION	AK026509.1
KEYWORDS	GI:10439384 Oligo capping; fls (full insert sequence). mouse carcinoma signet-ring cell carcinoma cell_line:KATO III CDNA t

## ORGANISM

NEW YORK

ARIKAWA, M., Yamashita, K., Kumagai, A., Itakura, S., Tanaka, T.,  
 Oga, T., Suzuki, Y., Obayashi, M., Nishii, T., Shiohara, T., Tanaka, T.,  
 Nakamura, Y., Isogai, T. and Sugano, S.  
 NEDO Human Genome Sequencing Project  
 (Funded by ¥2,000,000,000)  
 Grant # 8650589, Ministry of Science, Culture and Education  
 to S. Tanaka, T. Oga, T. Obayashi, M. Nishii, T. Isogai, T.,  
 Nakamura, Y. and Nakamura, Y.

**COMMENT**

## FEATURES

Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan  
(E-mail: cchaj@nids.n-tokyo.ac.jp, Tel: 81-3-5449-5286, Fax: 81-3-5449-5416)  
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; Department of construction, 5', & 3'-end one pass sequencing; Department of Medical Science, Virology and Human genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency). Location/Qualifiers

misc\_feature

North

4	BASE COUNT
	ORIGIN
1	100
2	100
3	100
4	100
5	100
6	100
7	100
8	100
9	100
10	100
11	100
12	100
13	100
14	100
15	100
16	100
17	100
18	100
19	100
20	100
21	100
22	100
23	100
24	100
25	100
26	100
27	100
28	100
29	100
30	100
31	100
32	100
33	100
34	100
35	100
36	100
37	100
38	100
39	100
40	100
41	100
42	100
43	100
44	100
45	100
46	100
47	100
48	100
49	100
50	100
51	100
52	100
53	100
54	100
55	100
56	100
57	100
58	100
59	100
60	100
61	100
62	100
63	100
64	100
65	100
66	100
67	100
68	100
69	100
70	100
71	100
72	100
73	100
74	100
75	100
76	100
77	100
78	100
79	100
80	100
81	100
82	100
83	100
84	100
85	100
86	100
87	100
88	100
89	100
90	100
91	100
92	100
93	100
94	100
95	100
96	100
97	100
98	100
99	100
100	100

Query Match	82.08;	Seed No. 3,7e+02;	Gaps	0;
Best Local Similarity	94.48;	Pred. 1		
Matches	17;	Conservative	0;	Mismatches 1;
QY	3	agatggtataagtagaac	20	
Db	222	ACATGCTATTAATAGAAC	239	

RESULT 15

HSCANLDA  
LOCUSDEFINITION  
ACCESSION

ACCESSION  
VERSION

**KEYWORDS**  
**SOURCE**

ORGANIS

112

REFERENCE  
AUTHORS

TITLE

JOURNAL

## REFERENCES

AUTHOR:
TITLE:

JOURNALS

## FEATURES

CDS

1

**This Page Blank (uspto)**

Fri Jun 8 09:43:27 2001

us-09-507-242-2.rag

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: June 8, 2001, 09:37:50 ; Search time 20.01 Seconds  
(without alignments)  
505.642 Million cell updates/sec

Title: us-09-507-242-2

Perfect score: 932  
Sequence: 1 MATSTGGLPPNMTIRVSRSH.....GEVSNITETSGVHLQRTG 177

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: A-Geneseg.0401:\*  
2: /SIDS6/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SIDS6/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SIDS6/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SIDS6/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SIDS6/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SIDS6/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SIDS6/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SIDS6/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SIDS6/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SIDS6/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SIDS6/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SIDS6/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SIDS6/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SIDS6/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SIDS6/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SIDS6/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SIDS6/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SIDS6/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SIDS6/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SIDS6/gcgdata/geneseq/geneseq/AA2000.DAT:\*  
22: /SIDS6/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	932	100.0	177	21	Y95876 Candida albicans C
2	369.5	39.6	163	18	W18312 NIMA-interacting p
3	313.5	33.6	163	20	Y48377 Human prostate can
4	303.5	32.6	119	21	G06627 Arabidopsis thalia
5	303.5	32.6	119	21	G14662 Arabidopsis thalia
6	303.5	32.6	119	21	G43433 Arabidopsis thalia
7	157	16.8	92	21	Y90945 Ceraurhacnem symbio
8	142.5	15.3	320	18	W89744 Staphylococcus aur
9	137	14.7	229	21	G08943 Arabidopsis thalia
10	137	14.7	221	21	G08942 Arabidopsis thalia
11	137	14.7	303	21	G08941 Arabidopsis thalia

12	127.5	13.7	142	21	G11000
13	127.5	13.7	169	21	G10999
14	121.5	13.0	204	20	W89814
15	121.5	13.0	299	18	W55450
16	121.5	13.0	299	19	W98322
17	121.5	13.0	299	19	W73034
18	121.5	13.0	299	20	W89829
19	121.5	13.0	299	20	W89849
20	121.5	13.0	39	21	B21976
21	117	12.6	317	17	R97692
22	116.5	12.5	317	20	Y19968
23	116.5	12.5	336	20	Y19968
24	102	10.9	131	19	W9203
25	101.5	10.8	427	20	Y08999
26	101.5	10.8	138	21	G00864
27	95.5	10.2	115	21	G34784
28	95.5	10.2	143	21	G34784
29	93	10.0	39	21	B21975
30	92	9.9	31	21	B21943
31	92	9.9	31	21	B21943
32	86.5	9.3	428	20	Y09000
33	85	9.1	1584	21	Y84430
34	85	9.1	1693	21	B48457
35	85	9.1	1713	21	B48457
36	85	9.1	1724	21	B48457
37	85	9.1	1724	21	B48457
38	84.5	9.0	877	19	W98214
39	84	9.0	331	21	G05720
40	84	9.0	638	21	G38502
41	84	9.0	720	21	G38501
42	84	9.0	777	21	G38500
43	83.5	8.9	286	18	W55712
44	82.5	8.8	62	8	P70035
45	82	8.8	326	22	B60765

#### ALIGNMENTS

RESULT 1	
Y95876	standard; Protein; 177 AA.
AC	Y95876;
XX	20-NOV-2000 (first entry)
XX	Candida albicans Caes1 proteins.
XX	Caes1; Ess1; infection; diagnosis; therapy; antifungal; fungicide;
XX	antiproliferative; cytostatic; antitumour.
KW	Candida albicans.
XX	WO200050561-A2.
XX	31-AUG-2000.
PD	18-FEB-2000; 2000MO-US04203.
PF	23-FEB-1999; 9905-0121246.
XX	(HEAL-) HEALTH RES INC.
PA	Hanes SD, Devasahayam G, Chaturvedi V;
PI	WPI: 2000-565453/52.
DR	N-PSDB; A50292.
XX	Novel Candida albicans gene, Caes1 useful for identifying compounds
PT	that specifically bind to and/or inhibit Caes1 and thus for treating
PT	Candida albicans infections and other life-threatening fungal
PT	infections -

Arabidopsis thalia  
Arabidopsis thalia  
Protein encoded by  
H. pylori ORF 02ae  
H. pylori GHPO 136  
Helicobacter pylori  
Protein encoded by  
Antigen from clust  
Ess1/9c peptide co  
Mouse Ess1 Wt doma  
B. burgdorferi ant  
Human parvulin-11k  
S. typhimurium sur  
Human secreted pro  
Arabidopsis thalia  
Arabidopsis thalia  
Ess1/S.c. peptide  
Pnl1/human peptide  
E. coli sura prote  
Human laminin 5 po  
Human laminin 5 po  
Human laminin 5 po  
H. pylori GHPO 108  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
H. pylori ORF 06gp  
Secretory signal s  
Gene 16 related pe

XX Claim 4; Fig 1A; Sipp; English.  
 PS This is the predicted translation product, Caes1, of the Caes1  
 XX gene from *Candida albicans*. Caes1 is the *C. albicans* homologue of  
 CC *Saccharomyces cerevisiae* ESS1, showing 428 amino acid identity.  
 CC Caes1 protein is a target for screening for antifungal and  
 CC inhibitor compounds useful for treating or preventing *C. albicans*  
 CC infections, and for raising antibodies useful for diagnostic  
 CC purposes or for blocking Caes1 activity. A Caes1 inhibitor can  
 CC be a compound which selectively inhibits growth of *S. cerevisiae*  
 CC not containing an endogenous ESS1 gene but rather Caes1 and  
 CC uninduced PIN1 (a human homologue of ESS1) and/or preferably does  
 CC not inhibit induced PIN1, e.g. does not inhibit *S. cerevisiae* not  
 CC containing an endogenous ESS1 but rather induced PIN1. Compositions  
 CC which inhibit PIN1 are useful antiproliferatives e.g. antineoplastics,  
 CC antitumor agents or anticancer agents. An anti-Caes1 antibody  
 CC or an antibody against an epitopic region of Caes1 can also be an  
 CC inhibitor of Caes1 by virtue of the antibody being able to bind to  
 CC Caes1.  
 XX  
 XX Sequence 177 AA:  
 SQ  
 Query Match 100.0%; Score 932; DB 21; Length 177;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-90;  
 Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MASSTGLPPMTIRVSRSHNKEFLNOSTNESSWDPPYGTDEKVLNAYIAKFRNNGKPLVNE  
 DB 1 masstglppmtirvrsrhshnkeyflnqstnesswdppytldkevlnaylakfrnngykp 60  
 QY 61 LVNEDGVVRVSHLLIKNNQSRKPKSWSPDGISRTDESIOLKKHLERILSGEVKLSSEL 120  
 DB 61 lvnedgvvrsvhlliknnqsrkpkswspdgisrtdesiolkkhlerrilsgvksel 120  
 QY 121 ANTESDCSSHDREGDLGFESKGMQMPPEEAFFNLHAGEVSNIIETNSGVHILORTG 177  
 DB 121 antesdcsshdregdlgffskgmppeeaafnlhgevsniietnsgvhlgrtg 177  
 RESULT 2  
 ID W18312 standard; Protein; 163 AA.  
 AC W18312;  
 XX 09-DEC-1997 (first entry)  
 DT NIMA-interacting protein Pin1.  
 XX NIMA-interacting protein Pin1.  
 DE  
 XX Pin1; protein interacting with NIMA; cell proliferation; mitosis;  
 KM peptidyl-propyl cis/trans isomerase; adenocarcinoma; cancer;  
 KM leukemias; psoriasis; pemphigus vulgaris; rheumatoid arthritis;  
 KM acute respiratory distress syndrome; septic shock; inflammation;  
 KM therapy.  
 OS  
 XX Homo sapiens.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 4..43  
 FT /label= MW  
 FT /note= "conserved tryptophan domain thought to  
 FT be mediate protein-protein interactions"  
 FT Peptide 54..69  
 FT /note= "putative nuclear localisation signal"  
 FT Domain 59..163  
 FT /label= PPI  
 FT /note= "peptidyl-propyl cis/trans isomerase domain"  
 XX MO9717986-A1.  
 XX 22-MAY-1997.  
 PD

XX 28-OCT-1996; 96MO-US17334.  
 XX  
 XX 13-NOV-1995; 95US-0555912.  
 XX  
 PA (SALK ) SALK INST BIOLOGICAL STUDIES.  
 XX  
 PI Hunter T, Lu KP;  
 XX  
 DR WPI: 1997-289057/26.  
 DR N-PSDB; T68888.  
 XX  
 XX Protein, Pin1, interacting with NIMA - used for treating cell  
 PT proliferative disorders  
 PS Claim 4; Page 49-50; 73pp; English.  
 CC This human polypeptide sequence comprises Pin1, an 18 kDa protein  
 CC that has peptidyl-propyl cis/trans isomerase activity, associates  
 CC with NIMA protein kinase, inhibits the mitosis-promoting function  
 CC of NIMA when overexpressed, and induces mitotic arrest and nuclear  
 CC fragmentation when depleted. Its sequence was deduced from a DNA  
 CC sequence (T68888) identified in an HeLa library using a yeast two-  
 CC hybrid system. A recombinant expression vector comprising the  
 CC DNA sequence and host cells containing the vector are claimed.  
 CC Methods are also claimed for identifying proteins that inhibit the  
 CC mitosis promoting function of NIMA protein kinase and for  
 CC controlling the growth of a cell by reducing Pin1 activity or PIN1  
 CC expression using an inhibitor, anti-Pin1 antibody, antisense  
 CC nucleotide sequence or ribozyme, or by increasing Pin1 activity  
 CC in the presence of an activator or increasing PIN1 expression using  
 CC an enhancer. This allows treatment of cell proliferation disorders  
 CC such as adenocarcinomas, cancers, psoriasis, pemphigus vulgaris, acute  
 CC respiratory distress syndrome, rheumatoid arthritis, septic shock  
 CC and inflammation.  
 XX  
 XX Sequence 163 AA:  
 SQ  
 Query Match 39.6%; Score 369.5; DB 18; Length 163;  
 Best Local Similarity 43.2%; Pred. No. 3.0e-31;  
 Matches 73; Conservative 34; Mismatches 49; Indels 13; Gaps 2;  
 QY 8 LPPMTIRVSRSHNKEFLNOSTNESSWDPPYGTDEKVLNAYIAKFRNNGKPLVNE  
 DB 7 lppmwkrmsrsgryvffnhitnasqwerpsgn-----ssgkngqgepar 54  
 QY 68 VRVSHLLIKNNQSRKPKSWSPDGISRTDESIOLKKHLERILSGEVKLSSELASDC 127  
 DB 55 vrcshllivkhsgrsrpswrg-ekltrkeaelingylqkiksgedfslasqfsc 113  
 QY 128 SSHRGDILGFESKGMQMPPEEAFFNLHAGEVSNIIETNSGVHILORT 176  
 DB 114 ssakargdlgrsqmqkpredastalrtgmsgpvftdsghililrt 162  
 RESULT 3  
 ID Y48377 standard; Protein; 163 AA.  
 AC Y48377;  
 XX 08-DEC-1999 (first entry)  
 DT Human prostate cancer-associated protein 74.  
 XX  
 DE  
 XX Human prostate cancer-associated protein 74.  
 XX Expressed sequence tag; EST; prostate; tumor; treatment; gene therapy;  
 KM cancer; tissue specificity; human.  
 XX  
 OS Homo sapiens.  
 OS  
 XX  
 XX Del9811194-A1.  
 XX  
 XX



PR	09-JUL-1999;	PR	99US-0142920.
PR	12-JUL-1999;	PR	99US-0142977.
PR	13-JUL-1999;	PR	99US-0143542.
PR	14-JUL-1999;	PR	99US-0143624.
PR	15-JUL-1999;	PR	99US-0144005.
PR	16-JUL-1999;	PR	99US-0144085.
PR	16-JUL-1999;	PR	99US-0144086.
PR	19-JUL-1999;	PR	99US-0144325.
PR	19-JUL-1999;	PR	99US-0144331.
PR	19-JUL-1999;	PR	99US-0144332.
PR	19-JUL-1999;	PR	99US-0144333.
PR	19-JUL-1999;	PR	99US-0144334.
PR	19-JUL-1999;	PR	99US-0144335.
PR	20-JUL-1999;	PR	99US-0144352.
PR	20-JUL-1999;	PR	99US-0144632.
PR	20-JUL-1999;	PR	99US-0144884.
PR	21-JUL-1999;	PR	99US-0144814.
PR	21-JUL-1999;	PR	99US-0145086.
PR	21-JUL-1999;	PR	99US-0145088.
PR	22-JUL-1999;	PR	99US-0145085.
PR	22-JUL-1999;	PR	99US-0145087.
PR	22-JUL-1999;	PR	99US-0145089.
PR	22-JUL-1999;	PR	99US-0145192.
PR	23-JUL-1999;	PR	99US-0145145.
PR	23-JUL-1999;	PR	99US-0145218.
PR	23-JUL-1999;	PR	99US-0145224.
PR	26-JUL-1999;	PR	99US-0145276.
PR	27-JUL-1999;	PR	99US-0145913.
PR	27-JUL-1999;	PR	99US-0145918.
PR	27-JUL-1999;	PR	99US-0145919.
PR	28-JUL-1999;	PR	99US-0145951.
PR	02-AUG-1999;	PR	99US-0146386.
PR	02-AUG-1999;	PR	99US-0146388.
PR	02-AUG-1999;	PR	99US-0146389.
PR	03-AUG-1999;	PR	99US-0146389.
PR	04-AUG-1999;	PR	99US-0147204.
PR	04-AUG-1999;	PR	99US-0147302.
PR	05-AUG-1999;	PR	99US-0147302.
PR	05-AUG-1999;	PR	99US-0147260.
PR	06-AUG-1999;	PR	99US-0147303.
PR	06-AUG-1999;	PR	99US-0147416.
PR	09-AUG-1999;	PR	99US-0147493.
PR	09-AUG-1999;	PR	99US-0147935.
PR	10-AUG-1999;	PR	99US-0148171.
PR	11-AUG-1999;	PR	99US-0148319.
PR	12-AUG-1999;	PR	99US-0148341.
PR	13-AUG-1999;	PR	99US-0148365.
PR	13-AUG-1999;	PR	99US-0148684.
PR	16-AUG-1999;	PR	99US-0149368.
PR	17-AUG-1999;	PR	99US-0149175.
PR	18-AUG-1999;	PR	99US-0149426.
PR	20-AUG-1999;	PR	99US-0149722.
PR	20-AUG-1999;	PR	99US-0149723.
PR	20-AUG-1999;	PR	99US-0149929.
PR	23-AUG-1999;	PR	99US-0149902.
PR	23-AUG-1999;	PR	99US-0149930.
PR	25-AUG-1999;	PR	99US-0150566.
PR	26-AUG-1999;	PR	99US-0150884.
PR	27-AUG-1999;	PR	99US-0151065.
PR	27-AUG-1999;	PR	99US-0151066.
PR	27-AUG-1999;	PR	99US-0151080.
PR	30-AUG-1999;	PR	99US-0151303.
PR	31-AUG-1999;	PR	99US-0151348.
PR	01-SEP-1999;	PR	99US-0151930.
PR	07-SEP-1999;	PR	99US-0152363.
PR	10-SEP-1999;	PR	99US-0153070.
PR	13-SEP-1999;	PR	99US-0153708.
PR	15-SEP-1999;	PR	99US-0154018.
PR	16-SEP-1999;	PR	99US-0154039.
PR	20-SEP-1999;	PR	99US-0154779.
PR	22-SEP-1999;	PR	99US-0155139.
PR	23-SEP-1999;	PR	99US-0155486.
PR	24-SEP-1999;	PR	99US-0155659.
PR	28-SEP-1999;	PR	99US-0156458.
PR	29-SEP-1999;	PR	99US-0156596.
PR	04-OCT-1999;	PR	99US-0157117.
PR	05-OCT-1999;	PR	99US-0157753.
PR	06-OCT-1999;	PR	99US-0157865.
PR	07-OCT-1999;	PR	99US-0158029.
PR	08-OCT-1999;	PR	99US-0158232.
PR	12-OCT-1999;	PR	99US-0158369.
PR	13-OCT-1999;	PR	99US-0159293.
PR	13-OCT-1999;	PR	99US-0159294.
PR	13-OCT-1999;	PR	99US-0159295.
PR	14-OCT-1999;	PR	99US-0159329.
PR	14-OCT-1999;	PR	99US-0159330.
PR	14-OCT-1999;	PR	99US-0159331.
PR	14-OCT-1999;	PR	99US-0159637.
PR	14-OCT-1999;	PR	99US-0159638.

PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130892.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132853.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 14-MAY-1999; 99US-0134768.  
PR 18-MAY-1999; 99US-0134941.  
PR 19-MAY-1999; 99US-0135124.  
PR 20-MAY-1999; 99US-0135353.  
PR 21-MAY-1999; 99US-0135629.  
PR 24-MAY-1999; 99US-0136392.  
PR 25-MAY-1999; 99US-0136392.  
PR 27-MAY-1999; 99US-0136782.  
PR 28-MAY-1999; 99US-0137282.  
PR 01-JUN-1999; 99US-0137528.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139817.  
PR 18-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139889.  
PR 22-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 23-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 29-JUN-1999; 99US-0141287.  
PR 30-JUN-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 26-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145818.  
PR 27-JUL-1999; 99US-0145919.  
PR 27-JUL-1999; 99US-0145951.  
PR 28-JUL-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 25-AUG-1999; 99US-0150884.  
PR 26-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151348.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0154018.  
PR 15-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154779.  
PR 20-SEP-1999; 99US-0155139.  
PR 22-SEP-1999; 99US-0155486.  
PR 23-SEP-1999; 99US-0155659.  
PR 24-SEP-1999; 99US-0156559.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.





PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	21-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	22-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	25-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	26-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	32.6%	Score 303.5	DB 21	Length 119
Best Local Similarity	51.3%	Pred. No. 2	2e-24	
Matches 58	Conservative 24	Mismatches 28	Indels 3	Gaps 1
OY	67	QVRVSHLLIKNNOSKPKPSKSPD---	ISRTDPSIGILKKHLERILSGEVLSELANI	123
Db	6	qvravshllkhgstrtkaswkdpegkllttr	eeaveqllsktrredlvsqkanfeevatr	65
OY	124	ESDSCSHDRGDLGFPSKGGOMPPPEEA	FNHVGEEVNIETNSGVHILQRT	176
Db	66	vsgcssektrgdlgsfgrgmqlptealt	yalvqldlvdtdsgvhlkrt	118
RESULT 7				
Y90945				
ID	Y90945	standard	Protein	92 AA
XX	Y90945			
XX	AC			
XX	DT			
XX	30-AUG-2000	(first entry)		
DE	Cenarchaeum symbiosum	open reading frame protein sequence	SFO	ID NO:68.
XX	Cenarchaeum symbiosum	non-thermophilic; crenarchaeote	physiology;	
KW	characterisation	archae; therapeutic; industrial;	laboratory.	
XX	OS			
XX	Cenarchaeum symbiosum.			
XX	PN	WO200018909-A2.		
XX	PD	06-APR-2000.		
XX	PF	29-SEP-1999;	99WO-US22752.	
XX	PR	29-SEP-1998;	98US-0102294.	
XX	PA	(DIVE-) DIVERSA CORP.		
XX	PI	Swanson RV, Feldman RA, Schleper C;		
XX	WPI	2000-293148/25.		
XX	N-PSDB	A55220.		

XX New nucleic acids and proteins isolated from the non-thermophilic  
PT crenarchaeote Cenarchaeum symbiosum, useful in characterizing the  
PT physiology of these archaea and in therapeutic, industrial or laboratory  
XX techniques -  
PS Claim 24; Page 186-187; 210pp; English.  
XX  
XX A55186 to A55226 and Y90913 to Y90951 represent nucleic acids and  
CC proteins isolated from the non-thermophilic crenarchaeote  
CC Cenarchaeum symbiosum. The nucleic acids and proteins identified in  
CC the present invention are useful in characterizing the physiology of  
CC these archaea and can be used in therapeutic, industrial or laboratory  
CC techniques. A55227 to A55260 represent promoter sequences from  
CC Cenarchaeum symbiosum. A55261 to A55269 represent PCR primers and  
CC probes used in examples from the present invention.  
XX  
SQ Sequence 92 AA;  
Query Match 16.8%; Score 157; DB 21; Length 92;  
Best Local Similarity 33.9%; Pred. No. 4.1e-09;  
Matches 38; Conservative 22; Mismatches 28; Indels 24; Gaps 4;  
OY 67 QVRVSHLIRKNNOSRKPSKSPDCISRTDPSIOILKHLERILSGEYKLSLANTES-125  
DB 4 kkschllvkk-----qgealav-----qerikage-kfkgklakeisl 40  
OY 126 DCSHHRGDLGFFSKGQMPPEEAFNLHVGEVSNIIETNSGVHIIQIRFG 177  
DB 41 dgsakrtgsglyfgfgrkmvrfedaatrlqygevsepvksefgyhvikrlg 92  
RESULT 8  
W89744 W89744 standard; Protein; 320 AA.  
AC W89744;  
DT 16-MAR-1999 (first entry)  
XX  
DE Staphylococcus aureus protein SEQ ID #5192.  
XX  
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome.  
XX  
OS Staphylococcus aureus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 17..32 /note="these residues represent a line of missing text  
FT in the sequence listing in the specification.  
FT They are included to maintain the residue  
FT numbering given in the specification for this  
FT protein sequence"  
XX  
PN EP786519-A2.  
XX  
PD 30-JUL-1997.  
XX  
PF 07-JAN-1997; 97EP-0100117.  
XX  
PR 05-JAN-1996; 96US-0009861.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
XX  
PI Rosen CA;  
XX  
DR WPI: 1997-374922/35.  
XX

PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -  
PT stored on computer readable medium and used in the production of  
PT anti-S.aureus vaccines  
XX  
PS Claim 23; Page 3187-3188; 3271pp; English.  
XX  
XX This sequence represents a Staphylococcus aureus protein sequence of the  
CC invention. The DNA sequences encoding the S.aureus proteins are recorded  
CC on a computer readable medium, preferably selected from a floppy or hard  
CC disk, random access memory (RAM), read-only memory (ROM) or CD-ROM.  
CC Homology searches using the S.aureus DNA sequences allows putative  
CC functions to be assigned so that protein-encoding or regulatory regions  
CC of commercial, therapeutic or industrial importance can be obtained.  
CC Specifically, sequences which are likely to encode antigens have been  
CC identified and these polypeptides can be used in a vaccine composition  
CC against S.aureus infection. The polypeptides can also be used in a kit  
CC for the immunodetection of S.aureus in a sample. S.aureus is implicated  
CC in numerous human diseases, including cellulitis, eyelid infections, food  
CC poisoning, osteomyelitis, skin and surgical wound infections, scalded  
CC skin syndrome, toxic shock syndrome, etc. Organisms transformed with the  
CC DNA sequences can be used for recombinant production of the polypeptides.  
CC The new DNA sequences (and their fragments) are useful as primers or  
CC probes for isolating homologues of any of the 5191 S.aureus DNA sequences  
CC contained on the computer readable medium.  
XX  
SQ Sequence 320 AA;  
Query Match 15.3%; Score 142.5; DB 18; Length 320;  
Best Local Similarity 33.3%; Pred. No. 8.5e-07;  
Matches 38; Conservative 23; Mismatches 44; Indels 9; Gaps 4;  
OY 62 VNEDGOVRVSHLIRKNNOSRKPSKSPDCISRTDPSIOILKHLERILSGEYKLSLA 121  
DB 137 lkeds-k-kashlllk-vksksdxeglddkaekqkaeeiq-----kvsdpkfkfgeia 188  
OY 122 NTES-DCSSHHRGDLGFFSKGQMPPEEAFNLHVGEVSNIIETNSGVHIIQ 174  
DB 189 kkesmdtgsakkdgelgyvlkgqtdfekalfklkgdevsevvksfgyhlik 242  
RESULT 9  
G08943 G08943 standard; Protein; 221 AA.  
ID G08943;  
AC G08943;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6677.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
XX

Fri Jun 8 09:43:27 2001

us-09-507-242-2.rag

PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136392.  
PR 27-MAY-1999; 99US-0136782.  
PR 28-MAY-1999; 99US-0137222.  
PR 01-JUN-1999; 99US-0137528.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137724.  
PR 07-JUN-1999; 99US-0138094.  
PR 08-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 10-JUN-1999; 99US-0139119.  
PR 14-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139899.  
PR 22-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142380.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145276.  
PR 26-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148341.  
PR 12-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 13-AUG-1999; 99US-0149368.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149422.  
PR 18-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0150060.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 01-SEP-1999; 99US-0152363.  
PR 07-SEP-1999; 99US-0153070.  
PR 10-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156596.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158368.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.





PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134766.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140699.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144633.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145066.  
  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148688.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154019.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.



PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158222.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.

PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160984.  
PR 25-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 13.7%; Score 127.5; DB 21; Length 142;  
Best Local Similarity 30.1%; Pred. No. 9.8e-06;  
Matches 34; Conservative 21; Mismatches 41; Indels 17; Gaps 2;  
  
OY 82 KPSKWSPDGTS-----RTDESIQILKKHLERLISGEVKL-----SELANTE 124  
Db 27 kykaagkaadgigtctcykarhnlcekgkineaykklidgwlsgdkvypaeafakiaey 86  
  
OY 125 SDCSSHDRGCDLGFPSKGOQPPFEPAFNLMWGEVSNIIETNSGVHILQRTG 177  
Db 87 secpsgkkgldgwfprjkmagpfgdvaInpvytsapfstbghyllseag 139  
  
RESULT 13  
ID G10999 standard; Protein; 169 AA.  
XX  
AC G10999;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 9548.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.



PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0144218.  
PR 14-MAY-1999; 99US-0144219.  
PR 14-MAY-1999; 99US-0144221.  
PR 14-MAY-1999; 99US-014370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139464.  
PR 18-JUN-1999; 99US-0139465.  
PR 21-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139899.  
PR 22-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 23-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140961.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 12-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144633.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148684.  
PR 13-AUG-1999; 99US-0149368.  
PR 16-AUG-1999; 99US-0149175.  
PR 17-AUG-1999; 99US-0149126.  
PR 18-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 15-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154779.  
PR 20-SEP-1999; 99US-0155139.  
PR 22-SEP-1999; 99US-0155486.  
PR 23-SEP-1999; 99US-0155659.  
PR 24-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157765.  
PR 06-OCT-1999; 99US-0158029.  
PR 07-OCT-1999; 99US-0158232.  
PR 08-OCT-1999; 99US-0158369.  
PR 12-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159285.  
PR 13-OCT-1999; 99US-0159285.  
PR 13-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 18-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160815.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.





...

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: June 8, 2001, 09:37:50 ; Search time 14.18 Seconds  
(Without alignments)  
857,825 Million cell updates/sec

Title: US-09-507-242-2

Perfect score: 932  
Sequence: 1 MASTSTGLPPNMTIRVSRSH.....GEVSNIIETNSGVHLIORTG 177

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: PIR-67:\*

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	409	43.9	175	T41093	peptidyl-prolyl ci
2	369.5	39.6	163	S68520	pini protein - hum
3	367	39.4	170	S52764	ESS1 protein - yeast
4	361.5	38.8	165	JC7136	peptidylprolyl iso
5	325	34.9	166	T08426	pini protein homol
6	170.5	18.3	292	S15269	post-translocation
7	166.5	17.9	333	A83797	protein secretion
8	142	15.2	221	T52622	probable peptidylp
9	124	13.3	621	A83412	peptidyl-prolyl ci
10	123	13.2	273	S52412	major antigenic pe
11	122	13.1	347	A81794	probable rotamase
12	122	13.1	348	B81216	peptidyl-prolyl ci
13	121.5	13.0	299	G64541	cell binding facto
14	121.5	13.0	299	B71967	probable peptidyl-
15	121	13.0	93	S48658	peptidylprolyl iso
16	119.5	12.8	92	A82396	peptidyl-prolyl ci
17	118.5	12.7	655	E82712	peptidyl-prolyl ci
18	117.5	12.6	282	E71662	protein export pro
19	116.5	12.5	336	D70113	basic membrane pro
20	116.5	12.5	619	E82141	peptidyl-prolyl ci
21	110.5	11.9	594	B64018	hypothetical prote
22	110.5	11.8	430	B83572	peptidyl-prolyl ci
23	109	11.7	93	C83123	peptidyl-prolyl ci
24	108	11.6	431	D82323	survival protein S
25	106	11.4	272	C83162	probable peptidyl-
26	105	11.3	292	F32055	nifm protein - Azo
27	104.5	11.2	92	F83395	peptidyl-prolyl ci
28	103	11.1	293	H43706	nifm-protein - Azo
29	97.5	10.5	297	S66102	protein secretion

30	95	10.2	126	2	T31601	hypothetical prote
31	89	9.5	428	2	B64726	probable peptidylp
32	88.5	9.5	314	2	B83559	probable peptidyl-
33	87.5	9.4	464	2	C82756	peptidyl-prolyl ci
34	85.5	9.2	879	2	E71811	probable iron-regu
35	85	9.1	1713	2	A55347	adhesive ligand ep
36	84.5	9.1	583	2	S30014	hypothetical prote
37	84.5	9.1	709	2	T16584	hypothetical prote
38	84.5	9.1	877	2	H64708	iron-regulated out
39	84	9.0	150	1	G69873	uncv-related prote
40	83.5	9.0	911	1	S51441	hypothetical prote
41	81.5	8.7	241	2	T25886	hypothetical prote
42	81	8.7	483	2	T19720	hypothetical prote
43	80.5	8.6	302	2	F64350	ribokinase (EC 2.7
44	80.5	8.6	468	2	T21331	hypothetical prote
45	80.5	8.6	1013	2	T46422	hypothetical prote

## ALIGNMENTS

RESULT 1  
T41093  
peptidyl-prolyl cis-trans isomerase - fission yeast (*Schizosaccharomyces pombe*)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_rev1500 03-Dec-1999 #text\_change 18-Aug-2000  
C:Accession: T41093  
R:Rumelle, B.; Goffeau, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, September 1998  
A:Reference number: 221964  
A:Accession: T41093  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-175 <PDB>  
A:Cross-references: EMBL:AL031535; PIDN:CA020742.1; GSPDB:GN00068; SPDB:SPCC16C4.03  
A:Experimental source: EMBL:Data Library, September 1998  
A:Gene: SPDB:SPCC16C4.03  
A:Map position: 3  
A:introns: 134/2  
A:Superfamily: yeast ESS1 protein; WW repeat homology  
F:4-42/Domain: WW repeat homology <WWR>

Query Match 43.9% Score 409; DB 2; Length 175;  
Best local similarity 46.0% Pred. No. 4.6e-28;  
Matches 80; Conservative 37; Mismatches 53; Indels 4; Gaps 4;  
QY 4 TSTGLPPNMTIRVSRSHNKEFLNOSTNESSWDPPYGTDEVLNAYIKFNNGYKPL-V 62  
DB 2 SNTGLKRPVIVKISRNRPYFNTETHSLWEPAAIDMALIKKFINELQESVTPTEA 61  
QY 63 NEDQVRVSHLTKNNOSRKPSPDCISRTROESTIOLKKHLERIL-SGEVKLSELA 121  
DB 62 SNPSKPIASHLTKHRSRSPSSWKE-EHITSKEEA-RKLAEHEDQLKSSVSMDLA 119  
QY 122 NTESDCSHRDGDLGFESKGQMQPPEEAFLHAGEVSNIIETNSGVHLIORTG 175  
DB 120 MKESDCSARRGGLGFEFGDEMQRPEDAALFKGEISGVETSSGPHIIR 173  
RESULT 2  
S68520  
Pini protein - human  
C:Species: Homo sapiens (man)  
C:Date: 15-Nov-1996 #sequence\_rev1500 27-Feb-1997 #text\_change 29-Sep-1999  
C:Accession: S68520  
R:Lin, K.P.; Hanes, S.D.; Hunter, T.  
Nature 380, 544-547, 1996  
A:Title: A human peptidyl-prolyl isomerase essential for regulation of mitosis.  
A:Reference number: S68520; MUID:96195064  
A:Accession: S68520  
A:Status: preliminary; nucleic acid sequence not shown

```
A:Molecule type: mRNA  
A:Residues: 1-163 <LUD>  
A:Cross-references: EMBL:U049070; NID:g1332709; PIDN:AAC50492.1; PID:g1332710  
C:Genetics:  
A:Gene: GDB:PIN1; dod  
A:Cross-references: GDB:521831  
C:Superfamily: Yeast ESS1 protein; MW repeat homology  
F,5-43/Domain: MW repeat homology <MWI>
```

Query Match            39,6%; Score 369,-5; DB 2; Length 163;  
Best Local Similarity     43,2% ; Pred. No, 1e-24;  
Matches      73; Conservative    34; Mismatches    49; Indels    13; Gaps        2.

OY    8 LPPNNITIRVSRSHNKKEYFLDNOSTNESSMDPYGTDEKVLAMAYIAFKKNNGKPLYNEDGQ 67  
     ||||| : | : | : | : | : | : | : | : | : | : | : | : |  
Db    7 LPFGCEHKRSRSGRGYYFFNHHTTNASOWERPGRN-----SSGGCKNGGGEIPAR 54

OY    68 VAVSHLLIKNNSRKRPKSMKSXPDGISRTROESIQILKKHLEKLISGEVKLSLATNESDC 127  
     |||||:::||||: || : |:|:::|:|:| :|:|:| |||  
Db    55 VRCSHLTVKHSSOSRRPSSWMRO-EKITRTKEALELINGYIOKIKSGEEDFEIASQFSDC 113

OY    128 SSSHGDGGLGFPSGOMOPPEEEAFNLHVGEVSNIETNSGVHLIRT 176  
     || ||||| ||||| ||||| : |||:| :|:||||| |||  
Db    114 SSARKRGDLGAFTSREGOMOKPFEDASFALTRTGEMSGPVFTDSGIHLILT 162

RESULT 3

552764

ESS1 protein - yeast (*Saccharomyces cerevisiae*)

N.Alternate names: protein J145; protein YJR017c; PFI1 protein

C.Species: *Saccharomyces cerevisiae*

C.Date: 19-May-1995 #sequence.revision 01-Sep-1995 #text\_change 29-Sep-1999

C.Accession: S52764; S55207; S57032; S07667; S65929

R.Hani, J.; Stumpf, G.; Domdey, H.

Submitted to the EMBL Data Library, March 1995

A.Description: PFI1 encodes an essential protein in *Saccharomyces cerevisiae*, which shows

A.Reference number: S52762

A.Accession: S52764

A.Molecule type: DNA

A.Residues: 1-170 <HAN>

A.Cross-references: EMBL:X85972; NID:q756283; PIDN:CAA59961.1; PID:q756286

R.de Haan, M.; Smits, P.H.M.; Grivell, L.A.

Submitted to the EMBL Data Library, May 1995

A.Reference number: S55183

A.Accession: S55207

A.Molecule type: DNA

A.Residues: 'MTIVVLPRTNRRYHNNDAP', 1-170 <DEN>

A.Cross-references: EMBL:X87611; NID:q854567; PIDN:CAA60941.1; PID:q854592

R.de Haan, M.; Grivell, L.A.; Smits, P.H.M.

Submitted to the Protein Sequence Database, September 1995

A.Reference number: S56771

A.Accession: S57032

A.Molecule type: DNA

A.Residues: 'MTIVVLPRTNRRYHNNDAP', 1-170 <ZAG>

A.Cross-references: EMBL:Z49517; NID:q1015651; PIDN:CAA89541.1; PID:q1015652; MIPS:YJR01

R.Hanes, S.D.; Shank, P.R.; Bostlian, K.A.

Yeast 5, 55-72, 1989

A.Title: Sequence and mutational analysis of ESS1, a gene essential for growth in *Saccha*

A.Reference number: S07667; MUID:89189095

A.Accession: S07667

A.Status: not compared with conceptual translation

A.Molecule type: DNA

A.Residues: 'MTIVVLPRTNRRYHNNDAP', 1-17, 'S', 9-16, 'A', 18-126, 'AASA', 133, 'S', 135-136, 'ARC',

A.Experimental source: strain DBY64

R.Hani, J.; Stumpf, G.; Domdey, H.

FEBS Lett. 365, 198-202, 1995

A.Title: PFI1 encodes an essential protein in *Saccharomyces cerevisiae*, which shows stric

A.Reference number: S65927; MUID:95300974

A.Accession: S65929

A.Molecule type: DNA

A.Residues: 1-170 <HAN>

A.Cross-references: EMBL:X85972; NID:q756283; PIDN:CAA59961.1; PID:q756286

C:Genetics:  
A:Gene: SGD:ESS1; PTF1  
A:Cross-references: SGD:S0003778; MIPS:YUR017c  
A:Map position: 10R  
C:Superfamily: Yeast ESS1 protein; MW repeat homology  
E:9-47/Domain: MW repeat homology <MW1>

Query Match	39.48;	Score 367;	DB 2;	Length 170;
Best Local Similarity	43.38;	Pred. No. 1.8e-24;		
Matches 77;	Conservative 35;	Mismatches 50;	Indels 16;	Gaps 4;

```

OY      3 STGSLGPNMNTIIVSRSSHNKCYEFPLNOSINNESMOPPGTGDDEVYLNATYAFKNGYRPLY 62
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      6 ASRGLPTPWIVRYRSKSKRRYEFNPNETKHSQWEEPGTNNQJL-----BKHLR 54

OY      63 NEDGQVRSHLLIINNOSRKPMSKSPDGISRTDDESIQILKHLERILSGEYKLSB--- 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      55 DHPYVRKRLHLILHKDSRRASHRS-ENITISKODATDELKTLITR-LDDSKTNSFEA 112

OY      120 LANTESDSSHDREGDLGFSGCOMPPPEEAFLNLHVGEYSNITLFNSGVHILQRTPG 177
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      113 LAKESDSSYKRGKGDIGWFERGEMQPSFEEDAAQLVGEYSVDLVEGSGGVHATIKRNG 170

```

RESULT 4

JC7136

peptidylprolyl isomerase (EC 5.2.1.8) - mouse

N:Alternate names: parvulins, Pim1 protein

C:Species: Mus musculus (house mouse)

C:Date: 04-Mar-2000 #sequence\_revision 04-Mar-2000 #text\_change 18-Aug-2000

C:Accession: JC7136

R:Fujimori, F.; Takahashi, K.; Uchida, C.; Uchida, T.

Biochem. Biophys. Res. Commun. 265, 658-663, 1999

A:Title: Mice lacking Pim1 develop normally, but are defective in entering cell cycle

A:Reference number: JC7136; MUID:20070807

A:Accession: JC7136

A:Molecule type: mRNA

A:Residues: 1-165 <PDB>

A:Cross-references: DDBJ:AB009691; NID:96468199; PIDD:BA087037.1; PIDD:96468200

C:Comment: This protein is a member of peptidyl prolyl cis/trans isomerase family. It is in oligopeptides, controls G2/M transition of HeLa cells, regulates entry into G1

C:Genetics:

A:Gene: Pim1

C:Superfamily: yeast Ess1 protein; WW repeat homology

C:Keywords: cis-trans-isomerase; mitosis

;-5-43/Domain: WW repeat homology <WNR>

	Query Match	38.8%	Score 361.5	DB 2;	Length 165;	
	Best Local Similarity	42.68%	Prod. No. 5e-24;			
	Matches 72; Conservative	35;	Mismatches 51;	Indels 11;	Gaps 3;	
OY	8 LPPMNTIVSRSHNKEVFLNSTNESSMDPPGTGTEKVELNAVIAFKNNGYKLNVEDQ	67				
	:           :           :					
Db	7 LPPEGEKMSRSGSVYYFNHITTNASQMERPSG-----GSIVGGSSKNKG----GEPAK	56				
OY	68 VRSLLILIKNNOSRKPKSKWKSFDGISRTIDESIQTLKKHLERILISGEVKLSLANTESDC	127				
	:           :           :           :					
Db	57 VRCSHLILVKSOSRRPSSWMRO-EKITRSKEALBELINGTIOKIKSGEEDFESIASOFSDC	115				
OY	128 SSHDGGGLGFPSKGOMOPPEEAFFNLHVGVSNIIETNSCVHTLRQ	176				
	:           :           :           :					
Db	116 SSAAKRGDLGPFRSGOMQPFEDASFAIRLTGMSPVPFDDSIHILRL	164				

RESULT 5  
T08426  
Pin1 protein homolog dodo - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*  
C:Date: 11-Jun-1999 #sequence-revision 11-Jun-1999 #text-change 17-Nov-2000  
C:Accession: T08426  
C:Maletzka, R.; de Couet, H.G.; Miklos, G.L.

[illegible]

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*  
A:Reference number: A69580; MUID:98044033  
A:Accession: H69682  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1292 <NUN>  
A:Cross-references: GB:J299109; GB:AL009126; NID:g2633260; PIDN:CAB12835.1; PID:e118229  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: prsA

[illegible]

RESULT 7  
A83797  
protein secretion (post-translocation chaparonin) prsa (imported) - *Bacillus halodurans*  
C:Species: *Bacillus halodurans*  
C:Date: 01-Dec-2000 #sequence-revision 01-Dec-2000 #ext-change 01-Dec-2000  
C:Accession: A83797  
R:Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., Fujii, F., & H  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* a  
A:Reference number: AB5650; MUID:2026334  
A:Accession: A83797  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-333 SSTS>  
A:Cross-references: GB:BA000004; NID:910173727; PIDN:BA04896.1; GSPDB:G  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: prsa

```

Query Match          17.9%; Score 166.5; DB 2; Length 333;
Best Local Similarity 30.1%; Pval: 6.4e-07;
Matches 50; Conservative 25; Mismatches 40; Indels 51; Gaps 7;

QY      21 NKEYVLTNGSTNEMSDPPYGTDR-----EVLNRYIAKFKNNQIKPLVWEDG--QV 68
      ||| |||:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      114 NKEEFVNYIL-----PPLVLEKLAVEGVETDEKQKVPDE-----NRDSLTIEY 158
      ||| |||:| | | | | | | | | | | | | | | | | | | | | | | | | |
QY      69 RYSHLLIKNNQSRPKSKWSPDQISTROESIQILKKHLERLISGEVKLSELTATES-DC 127
      |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      159 EASHILVEDEFT-----AEVLDRLRENG--DFEALASEYVDP 195
      |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY      128 SSMDRGGDGLGFSPKQMPPEEAAFNILHAGEVSNIIETNSGVHL 173
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      196 SAEANNNGDLGFPGKQDMVPPEFEAAFMELDEIVSEPESTGYHIT 241
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT      8
T52622
probable peptidylprolyl isomerase (EC 5.2.1.8) [imported] - Arabidopsis thaliana
N:Alternate names: Probable peptidyl-prolyl cis-trans isomerase
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000
C:Accession: T52622

```

R:Babyhuk, E.; Fuangthong, M.; Van Montagu, M.; Inze, D.; Kushnir, S.  
 Proc. Natl. Acad. Sci. U.S.A. 94, 12722-12727, 1999  
 A:Title: Efficient gene tapping in Arabidopsis thaliana using a gene trap approach.  
 A:Reference number: 226135  
 A:Accession: T52622  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-221 <BAB>  
 A:Cross-references: EMBL:286095; PIDN:CAB06699.1  
 A:Experimental source: cultivar Columbia  
 C:Keywords: cis-trans-isomerase

Query Match 15.2%; Score 142; DB 2; Length 221;  
 Best Local Similarity 33.3%; Pred. No. 4.7e-05;  
 Matches 36; Conservative 17; Mismatches 33; Indels 22; Gaps 3;

Oy 67 GVRVSHLLIKNNQSKRPSKSPDISRTDESIOILKKHLERILSGEVKSELANTESD 126  
 Db 17 ELVQHLVKKNN-----DVELFAELQKKFLDGE-EMSDLAAYSI 55

Oy 127 CSSHNRGDLGFFSGKQMPPEEAFNLHGVESNIETNSGVHILQ 174  
 Db 56 CPSSKKDGLGWKVGKGVVPEEEAFKAEINQVVR-CRQPGHLLQ 102

RESULT 9  
 A83421  
 pepTidY1-prolyl cis-trans isomerase D PA1805 [imported] - Pseudomonas aeruginosa (strain  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: A83421  
 R:Stover, C.K.; Plam, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laidig, K.; Lim,  
 J.; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: A83421  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-621 <STO>  
 A:Cross-references: GB:A8004606; GB:A8004091; NID:g9947780; PIDN:AA05194.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:

A:Gene: ppid; PA1805

Query Match 13.3%; Score 124; DB 2; Length 621;  
 Best Local Similarity 25.7%; Pred. No. 0.0064;  
 Matches 39; Conservative 26; Mismatches 59; Indels 28; Gaps 3;

Oy 41 TDEVLNAYIAKFNKNNGY-----KPLVNDGQVRSVSHLLIKNNQSKRPS 85  
 Db 223 TPEQVVVEVELKSSFPDQKVKQEDLEALYKQRIANLSEGRDAHLITIEVND----- 276

Oy 86 WSPDGISRTDESIOILKKHLERILSGEVKSELANTESDSSHDRGDLGFFSGKQMQ 145  
 Db 277 -KVGDEQAKAKIDEIK-----ARLAKGEDFAALAKESQDGSAAATGDDGAGRGVYD 329

Oy 146 PPEEAFNLHGVESNIETNSGVHILQRTG 177  
 Db 330 PAFEEALYALKQGEVSAPVKTPTGYHLITLIG 361

RESULT 10  
 S52412  
 major antigenic peptide PEB3/cell binding factor 2 Cj0596 precursor [imported] - Campylo  
 N:Alternate names: 31k antigen PEB4  
 C:Species: Campylobacter jejuni  
 C:Date: 08-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 05-May-2000  
 C:Accession: S52412; A60701; D41161; C81407  
 R:Burucoa, C.; Fremaux, C.; Pel, Z.; Murali, T.; Blaser, M.; Cenatiempo, Y.; Fauchere, J

submitted to the EMBL Data Library, February 1995  
 A:Description: Cloning and sequencing of Cbf2 and external protein from Campylobacter  
 A:Reference number: S52412  
 A:Accession: S52412  
 A:Molecule type: DNA  
 A:Residues: 1-273 <BUR>  
 A:Cross-references: EMBL:X84703; NID:9671839; PIDN:CAAS9175.1; PID:9671840  
 R:Dubreuil, J.D.; Kostzyńska, M.; Logan, S.M.; Harris, L.A.; Austin, J.W.; Trust, T.  
 J. Clin. Microbiol. 28, 1321-1328, 1990  
 A:Title: Purification, characterization, and localization of a protein antigen shared  
 A:Reference number: A60701; MUID:90338381  
 A:Accession: A60701  
 A:Molecule type: Protein  
 A:Residues: 22-34, 'A', 36-54, 'X', 56-58, 'A' <DU2>  
 A:Experimental source: strain VC74  
 A:Note: this protein was purified after acid extraction from the surface of intact ba  
 R:Pel, Z.; Ellison III, R.T.; Blaser, M.J.  
 J. Biol. Chem. 266, 16363-16369, 1991  
 A:Title: Identification, purification, and characterization of major antigenic prot  
 A:Reference number: A41161; MUID:91358413  
 A:Accession: D41161  
 A:Molecule type: Protein  
 A:Residues: 22-34, 'A', 36-51, 'X', 53, 'X', 55 <PEI>  
 A:Experimental source: strain 81-176  
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churcher, C.; Basham, D.; Chil  
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Ba  
 Nature 403, 665-668, 2000  
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals  
 A:Reference number: A81250; MUID:20150912  
 A:Accession: C81407  
 A:Molecule type: DNA  
 A:Residues: 1-273 <PAR>  
 A:Cross-references: GB:AL139075; GB:AL111168; NID:96967817; PIDN:CAB75232.1; PID:9696  
 A:Experimental source: serotype O2, strain NCTC 11168  
 C:Genetics:

A:Gene: cbp4/cbf2; Cj0596  
 C:Keywords: cell binding; membrane-associated protein  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-273/Product: cell binding factor 2 #status experimental <MAT>

Query Match 13.2%; Score 123; DB 2; Length 273;  
 Best Local Similarity 28.5%; Pred. No. 0.0026;  
 Matches 39; Conservative 24; Mismatches 44; Indels 30; Gaps 6;

Oy 42 DKEVLNAYIAKFNKNNGYPLVNDGQVRSVSHLLIKNNQSKRPSDGISRTDESIO 101  
 Db 113 DAAKKAFTYDQNKDKYKPF-----ARVQAKHIVATEKEAK-----DIIN 152

Oy 102 ILKKHLERILSG---EVKLSLANTES-DCSSHNRGDLGFFSGKQMPPEEAFNLHV 157  
 Db 153 ELKG-----LKGKELDAFSELAKEKSIDPSKNGGELGWFDSQVWKPPTDAFAALKN 207

Oy 158 GEVSNL-ETNSGVHIL 173  
 Db 208 GTTTTTPVKTNGYHVI 224

RESULT 11  
 A81794  
 probable rotamase NMA2206 [imported] - Neisseria meningitidis (group A strain Z2491)  
 C:Species: Neisseria meningitidis  
 C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 05-May-2000  
 C:Accession: A81794  
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo  
 Hoiroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre  
 Nature 404, 502-506, 2000  
 A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491  
 A:Reference number: A81775; MUID:20222556  
 A:Accession: A81794  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-347 <PAR>





R:Rahfeld, J.; Ruecknagel, K.P.; Schelbert, B.; Ludwig, B.; Hacker, J.; Mann, K.; Fische  
 A:Submitted to the Protein Sequence Database, August 1994  
 A:Description: Confirmation of the existence of a third family among peptidyl-prolyl cis  
 A:Reference number: S45525  
 A:Accession: S45525  
 A:Molecule type: protein  
 A:Residues: 2-93 <RAH1>  
 A:Experimental source: strain K-12  
 R:Rahfeld, J.U.; Schlerhorn, A.; Mann, K.; Fischer, G.  
 FEBS Lett. 343, 65-69, 1994  
 A>Title: A novel peptidyl-prolyl cis/trans isomerase from Escherichia coli.  
 A:Reference number: S43654; MUID:94215709  
 A:Accession: S43654  
 A:Molecule type: protein  
 A:Residues: 2-22 <RAH2>  
 A:Experimental source: strain K-12  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC  
 .A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A>Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: B65181  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-93 <BLAT>  
 A:Cross-references: GB:AE00454; GB:U00096; MID:92367278; PIDN:AAC76780.1; PID:91790211;  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Genetics:  
 A:Gene: PplC  
 C:Keywords: cis-trans-isomerase; cyclosporin A binding

Query Match 13.0%; Score 121; DB 2; Length 93;  
 Best Local Similarity 37.3%; Pred.No. 0.00096;  
 Matches 31; Conservative 13; Mismatches 31; Indels 8; Gaps 2;  
 QY 99 STQILKKH-----LERLSGEVKLSLEANTESDCSSHDRCGDLGFRSKGOMPPFEPA 151  
 Db 7 ALHILVKEEKALDLLEQIKNG-ADFGKLAKKHSICPSGKRGDLEGFROGOMVPADFQV 65  
 QY 152 AFNLHVGESVNIETNSGVHILQ 174  
 Db 66 VESCPVLEPTGPIHTQFGYHITK 88

Search completed: June 8, 2001, 09:38:11  
 Job time: 21 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2001, 09:37:55 ; Search time 9.54 Seconds

(without alignments)  
635.558 Million cell updates/sec

Title: US-09-507-242-2

Perfect score: 932

Sequence: 1 MASTSTGLPPNMTIRVSRSH.....GEVSNIIETNSGVHILQRTG 177

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	369.5	39.6	163	1	PIN1_HUMAN
2	367	39.4	190	1	ESS1_YEAST
3	335	34.9	166	1	DOD_DROME
4	170.5	18.3	292	1	PRSA_BACSU
5	123	13.2	273	1	CBE2_CAMJE
6	121.5	13.0	299	1	Y175_HELPJ
7	121.5	13.0	299	1	Y175_HELPJ
8	121	13.0	299	1	Y175_HELPJ
9	114	12.2	100	1	PIN1_HUMAN
10	110.5	11.9	622	1	PIN1_HUMAN
11	105	11.3	292	1	NIFM_AZOVI
12	103	11.1	293	1	NIFM_AZOVI
13	97.5	10.5	297	1	YACD_BACSU
14	97	10.4	577	1	BAG3_MOUSE
15	89	9.5	428	1	SURA_ECOLI
16	85	9.1	113	1	LMA3_HUMAN
17	84.5	9.1	583	1	PR40_YEAST
18	81.5	8.7	1914	1	KMLS_HUMAN
19	80.5	8.6	302	1	Y406_METRA
20	80.5	8.6	421	1	GAS7_MOUSE
21	80.5	8.6	746	1	EXT1_HUMAN
22	80.5	8.6	918	1	YAMB_CAEEL
23	79.5	8.5	957	1	NED4_MOUSE
24	79.5	8.5	430	1	SURA_BUCAT
25	79	8.5	1814	1	TSC2_MOUSE
26	78.5	8.4	426	1	NIFM_KLEPN
27	78.5	8.4	422	1	GAS7_RAT
28	78.5	8.4	1664	1	RPA1_YEAST
29	78.5	8.4	2259	1	CCAE_RABIT
30	78.5	8.4	256	1	M3R_DICDI
31	78	8.4	313	1	SURA_HAEIN
32	78	8.4	387	1	UPF3_YEAST
33	78	8.4	387	1	UPF3_YEAST

34	77.5	8.3	724	1	Y061_CAEEL
35	77	8.3	259	1	KIK9_RAT
36	77	8.3	299	1	PRM1_LACPA
37	77	8.3	412	1	GAS7_HUMAN
38	76	8.2	549	1	PP01_YEAST
39	75.5	8.1	182	1	HSLV_BORBU
40	75.5	8.1	361	1	WDS_DROME
41	75	8.0	259	1	KIK2_RAT
42	75	8.0	955	1	MMU3_MYCLE
43	75	8.0	1022	1	P164_HUMAN
44	74.5	8.0	228	1	YDZ1_SCHPO
45	74.5	8.0	652	1	MX1_RAT

## ALIGNMENTS

RESULT 1  
ID PIN1\_HUMAN STANDARD; . PRT; 163 AA.  
AC Q13526;  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE NIMA-INTERACTING 1 (EC 5.2.1.8).  
GN PIN1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TAXID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE:96195064; Pubmed-8606777;  
RA Lu K.P., Hanes S.D., Hunter T.;  
RT "A human peptidyl-prolyl isomerase essential for regulation of  
mitosis.";  
RL Nature 380:544-547(1996)  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (1.35 ANGSTROMS).  
RX MEDLINE:97344079; Pubmed-9200606;  
RA Ranganathan R., Lu K.P., Hunter T., Noel J.P.;  
RT "Structural and functional analysis of the mitotic rotamase Pin1  
suggests substrate recognition is phosphorylation dependent.";  
RL Cell 89:875-886(1997).  
CC -1- FUNCTION: ESSENTIAL PPIASE THAT REGULATES MITOSIS PRESUMABLY BY  
INTERACTING WITH NIMA AND ATTENUATING ITS MITOSIS-PROMOTING  
ACTIVITY.  
CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC  
PEPTIDE BONDS IN OLIGOPEPTIDES.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- SIMILARITY: BELONGS TO THE PPI/PAVULIN FAMILY OF ROTAMASES.  
CC -1- SIMILARITY: CONTAINS 1 WW DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement. (See <http://www.isb.ch/announce/>  
or send an email to [license@isb.slb.ch](mailto:license@isb.slb.ch).)  
CC EMBL: U49070; AAC50492.1; -  
CC PDB: 1PIN; 25-NOV-98.  
CC MIM: 601052; -  
DR InterPro: IPR000297; -  
DR InterPro: IPR001202; -  
DR InterPro: IPR002349; -  
DR Pfam: PF00639; Rotamase; 1.  
DR Pfam: PF00397; WW; 1.  
DR PRINTS: PR00403; WMDOMAIN.  
DR PROSITE: PS01096; PPI\_C\_PPIASE\_1; 1.  
DR PROSITE: PS0198; PPI\_C\_PPIASE\_2; 1.  
DR PROSITE: PS01159; WW\_DOMAIN\_1; 1.

Df	PROSITE	PS0020; MW_DOMAIN_2; 1
KM	Isomerase; Rotamase;	Nuclear protein; Cell cycle; 3D-structure.
FT	DOMAIN	5 39 MW.
FT	DOMAIN	52 163 PIC-LIKE.
SQ	SEQUENCE	163 AA; 18243 MW; 35391AFA0B7D1E13 CRC64;
Query Match		
Best Local Similarity 39.6%; Score 369.5; DB 1; Length 163; Matched 73; Conservativity 43.2%; Pred. No. 1.6e-25; Matches 73; Conservativity 34; Mismatches 49; Indels 13; Gaps 2;		
OY	8	LPPNNTIRSRSHNKREYFLNSTSSNDPPTGCTKEVLAVIAKFNKNKYKPLVNEOG 67
Db	7	LPPGEKMRMSRSGAVYFNFHTITNASOMERBSGN-----SSGGKNGGEPAR 54
OY	68	VRVSHLLIKNNOSRRPKSMKSPDGISRTDESIQILKHLEIILSGEVKLSELANTESDC 127
Db	55	VYCSHLIVKHSQSRRPSSWRQ-EKITPRKEALTELINGYIQIKSGEDPFESLASQFSDC 113
OY	128	SSHDSGDLGFPSKCOMOPPEEAFLNLHGVSNIETNSGVHILQRT 176
Db	114	SSAKARGDLGAFSKGOMPPEDASFAIRTGTMSGSPVFSDSIIHLILT 162

RESULT	2	
ESS1_YEAST		
ID	ESS1_YEAST	STANDARD: PRT: 190 AA.
AC	P22696;	
DT	01-AUG-1991 (Rel. 19, Created)	
DT	01-NOV-1995 (Rel. 32, Last sequence update)	
DT	01-OCT-2000 (Rel. 40, Last annotation update)	
DE	ESS1 PROTEIN (PROCESSING/TERMINATION FACTOR 1).	
GN	ESS1 OR PPE1 OR YJR017C OR J1452.	
OS	Saccharomyces cerevisiae (Baker's yeast).	
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;	
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.	
OX	NCBI_TaxID=4932;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=89189095; PubMed=2648698;	
RA	Hanes S.D., Shank P.R., Bostian K.A.;	
RT	"Sequence and mutational analysis of ESS1, a gene essential for	
RL	growth in Saccharomyces cerevisiae.";	
RL	Yeast 5:55-72(1989).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=DH484;	
RX	MEDLINE=95300974; PubMed=7781779;	
RA	Hant J., Stumpf G., Domdey H.;	
RT	"Ppf1 encodes an essential protein in Saccharomyces cerevisiae, which	
RL	shows strong homology with a new putative family of Pifases.";	
RL	FEBS Lett. 365:198-202(1995).	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=SS288C / FY1679;	
RL	de Haan M., Smits P.H.M., Grievell L.A.;	
CC	Submitted (May-1995) to the EMBL/GenBank/DBJ databases.	
CC	- FUNCTION: ESSENTIAL FOR GROWTH IN SACCCHAROMYCES CEREVISIAE. MAY BE	
CC	INVOLVED IN CYTOKINESIS OR IN CELL SEPARATION.	
CC	- SIMILARITY: CONTAINS 1 MW DOMAIN.	
CC	- SIMILARITY: BELONGS TO THE PIFC/PARYULIN FAMILY OF ROTAMASES.	
CC	- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-21 IS THE INITIATOR.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
EMBL	X85972; CAA59961.1; ALT_INIT.	
EMBL	X87611; CAA60941.1; -	

DR	EMBL; Z49517; CA89541.1; -.
DR	PIR; S07867; S07867.
DR	SCD; S0003778; ESSL.
DR	InterPro; IPR000297; -.
DR	InterPro; IPR001202; -.
DR	InterPro; IPR002349; -.
DR	Pfam; PF00639; Rotamase; 1.
DR	Pfam; PF00397; WW; 1.
DR	PRINTS; PR00403; WMDOMAIN.
DR	PROSITE; PS01096; PPIC_PPIASE_1; 1.
DR	PROSITE; PS0198; PPIC_PPIASE_2; 1.
DR	PROSITE; PS01159; WW_DOMAIN_1; 1.
DR	PROSITE; PS0020; WW_DOMAIN_2; 1.
KW	Isomerase; Rotamase.
FT	DOMAIN 29 62
FT	DOMAIN 77 190
FT	CONFLICT 28 28
FT	CONFLICT 37 37
FT	CONFLICT 147 190
FT	
SEQUENCE	190 AA; 2173 MW; 93E449E3D7B4D989 CRC64; REF. 1).

```

Query Match          39.4%; Score 367; DB 1; Length 190;
Best Local Similarity 43.3%; Pred. No.3,le25;
Matches 77; Conservative 35; Mismatches 50; Indels 16; Gaps
4;

OY 3 STSTGLPPMTITRVSRSNHNKXYELNQSTNESSMDPPYGTDKKVLNAYIAKFKNNNGKPLV 62
    ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 26 ASRTGLPPMTIVRYSKSKKREYFNPETRKHSQWMEPEGINKDQL-----HKHLR 74
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 63 NEDGQVRSHLLIKNNQSKPKSKSPDCISITRDESIOLKKHLERLISGEVKLSE--- 119
    :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 75 DHPVVRCHLLILIKRHSRPPASHS-ENITTSKQDADELTLTLTR-LDDSKTNSFEA 132
    :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 120 IANFSDSSHDHRCDDLFEFSKGMQPPFEAAFLNHLGVSNIETNSGVHILORTG 177
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 133 LAKESSDCSSYKRGDGLMGFGKGEWQSPEDAAFLQKLGVSVDIYSSGSGVHIVRG 190
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 3
DOD_DROME STANDARD; PRT; 166 AA.
ID DOD_DROME
AC P54353; 061344; 09VRH1;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DODO PROTEIN.
GN DODO OR CG17051.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
OK
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=OREGON-R; TISSUE=Pupae, and Larva;
RC MEDLINE=96133954; PubMed=8552658;
RA Maleszka R., Hanes S.D., Hackett R.L., de Couet H.G., Miklos G.L.G.;
RT "The drosophila melanogaster dodo (dod) gene, conserved in humans, is
RT functionally interchangeable with the ESSI cell division gene of
RT Saccharomyces cerevisiae."
RT Proc. Natl. Acad. Sci. U.S.A. 93:447-451(1996).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=CANTON-S;
RC MEDLINE=98188272; PubMed=9520435;
RA Maleszka R., de Couet H.G., Miklos G.L.G.;
RT "Data transferability from model organisms to human beings: insights
RT from the functional genomics of the flyhthless region of Drosophila.";
RT Proc. Natl. Acad. Sci. U.S.A. 95:3731-3736(1998).
RN [3]

```

RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RA MEDLINE=20196006; PubMed-10731132;  
 RA Adams M.D., Ceinlker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,  
 RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.D., Evangelista C.C., Ferraz C., Ferrelia S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleib J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spletter E., Sprelling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Wooley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 CC -1- SIMILARITY: BELONGS TO THE PRIC/PARVULIN FAMILY OF ROTAMASES.  
 CC -1- SIMILARITY: CONTAINS 1 MW DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC  
 CC EMBL: U35140; AAC46958.1; -;  
 DR EMBL: AF017777; AAC28408.1; -;  
 DR EMBL: AE003568; AAF50829.1; -;  
 DR HSSP: Q13526; 1PIN.  
 DR PDBase: FBgn0015379; dod.  
 DR InterPro: IPR000297; -;  
 DR InterPro: IPR001202; -;  
 DR InterPro: IPR002349; -;  
 DR Pfam: PF00639; Rotamase; 1.  
 DR Pfam: PF00397; MW; 1.  
 DR PRINTS: PR00403; WNDOMAIN.  
 DR PROSITE: PS01096; PRIC\_PPIASE\_1; 1.  
 DR PROSITE: PS01198; PRIC\_PPIASE\_2; 1.  
 DR PROSITE: PS01159; MW\_DOMAIN\_1; 1.  
 DR PROSITE: PS00202; MW\_DOMAIN\_2; 1.  
 KW Isomerase; Rotamase.  
 FT DOMAIN 5 39 MW.  
 FT DOMAIN 55 166 PRIC-LIKE.  
 FT CONFLICT 44 166 A -> T (IN REF. 2).  
 FT SEQUENCE 166 AA; 18376 MW; 3B4306FA930E7259 CRC64;

Query Match 34.9%; Score 325; DB 1; Length 166;  
 Best Local Similarity 41.7%; Pred. No. 1,2e-21;  
 Matches 70; Conservative 30; Mismatches 58; Indels 10; Gaps 3;  
 QY 8 LPPNWTIRVSRINKEKFLPNQSTNESSWDPYGTQKEVLNATYAKKNNGYKPLVVEDQ 67  
 DB 7 LPDGMKRTSRSGMYLYLMYTKRESQMP--TEP-----AKKAGGSGAGGAPDE 57  
 QY 68 VRSHLLINKNOSRKPSKSMKSPDGISTRDESIQIKKHLERITLGEVKLSLANTESDC 127  
 DB 58 VHCILVLVHKHSGRRSSWREAN-ITRIKEAQLLEVYRANKIVQGEAFPFDELARSYDC 116  
 QY 128 SSHDRGSGDGFESKGMOPPEEAFNLHGEVSNITETNSGVHILQR 175  
 DB 117 SSKRRGGDGLKFGFRGMQAFAEDAAFLKLVNQLSGIVDSGGLHILR 164  
 RESULT 4  
 ID PRSA\_BACSU STANDARD; PRT; 292 AA.  
 AC P24327;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PROTEIN EXPORT PROTEIN PRSA PRECURSOR.  
 GN PRSA.  
 OS *Bacillus subtilis*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=92065824; PubMed=1956302;  
 RA Kontinen V.P., Saris P., Sarvas M.;  
 RT "A gene (prsa) of *Bacillus subtilis* involved in a novel, late stage  
 RT of protein export.";  
 RL Mol. Microbiol. 5:1273-1283(1991).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Noback M.A., Terpsstra P., Holsappel S., Venema G., Bron S.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: INVOLVED IN A LATE STAGE OF PROTEIN EXPORT. PROBABLE  
 CC PRASE.  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID  
 CC ANCHOR (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE PRIC/PARVULIN FAMILY OF ROTAMASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC  
 CC EMBL: X57271; CAA40543.1; -;  
 DR EMBL: Y14077; CAA74418.1; -;  
 DR EMBL: Z99109; CAB12835.1; -;  
 DR PIR: S15269; S15269.  
 DR Subtilist; BG10464; prsa.  
 DR InterPro: IPR000297; -;  
 DR Pfam: PF00639; Rotamase; 1.  
 DR PROSITE: PS01096; PRIC\_PPIASE\_1; 1.  
 DR PROSITE: PS01198; PRIC\_PPIASE\_2; 1.  
 DR PROSITE: PS00013; PROKAR\_LIPOPROTEIN; 1.  
 KW Isomerase; Rotamase; Membrane; Lipoprotein; Signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 292 PROTEIN EXPORT PROTEIN PRSA.  
 FT LIPID 20 20 N-ACYL DIGLYCERIDE (POTENTIAL).  
 FT DOMAIN 134 224 PRIC-LIKE.

```

SQ      SEQUENCE   292 AA;    32510 MW;    DEC194D8FE5C9FC0 CAC64;

Query Match          18.3%; Score 170.5; DB 1; Length 292;
Best Local Similarity 30.5%; Pred. No. 7.3e-08;
Matches 51; Conservative 29; Mismatches 46; Indels 41; Gaps 7;

OY      16 VSRSHNKREYFLNQ-----STNESSDDPYGCDKEVLNVIKFNKNNGVKPLVNEGCQVRV 70
DB      92 LERQIGKDTLEEQVAYELLTQNAADNIKVTADADIKETW----EGKK-----GKTIRA 139
OY      71 SHLLIKNNOSRKPKRSWKSPDGISRTDRSDSIQLTKHLEIRLISEGYKLSELANTESDCS8 130
DB      140 SHILVAD-----KKTAEEVEKKLRK-----GE-KPFDIAKEYSTDSSA 176
OY      131 DRGGDLGFFSK-GQMOPPEEAFAFNLHVGEVSNIETNSGHITLT 176
DB      177 SKGGDLGWFAKEGOMDETFSKAFRLTKGEVSDPKTYGVHIHKRT. 223

RESULT      5
CBF2_CAMJE ID CBF2_CAMJE STANDARD: PRT; 273 AA.
AC Q46105;
DF 01-NOV-1997 (Rel. 35, Created)
DF 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE CELL BINDING FACTOR 2 PRECURSOR (MAJOR ANTIGEN PEB4A).
GN CBF2 OR PEB4A OR CJ0596.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN RP
RC SEQUENCE FROM N.A.
RC STRAIN-ATCC 55026 / 81-176;
RX MEDLINE=96099687; PubMed=8525063;
RA Burucoo C., Fremaux C., Pel Z., Tummurru M., Blaser M.J.,
RA Genatlampo Y., Fauchere J.L.;
RT "Nucleotide sequence and characterization of peba4 encoding an
RT antigenic protein in Campylobacter jejuni.";
RL Res. Microbiol. 146:467-476(1995).
[2]
RP RE
RC SEQUENCE FROM N.A.
RC STRAIN-NCTC 11166;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Baeham D., Chillingworth T., Davies R.M., Felwell T., Holroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vleet A.H.M.,
RA Whitehead S., Barrrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
-1 SIMILARITY: BELONGS TO THE PRIC/PARVULIN FAMILY OF ROTAMASES.
STRONG TO H.PYLORI HP0175.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on way
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; X84703; CAAS9175.1; -.
CC EMBL; AL139075; CAB75232.1; -.
CC InterPro; IPR000297; -.
CC Pfam; PF00639; Rotamase_1.
CC PROSITE; PS01096; PRIC_PRIASE_1; 1.
CC PROSITE; PS01098; PRIC_PRIASE_2; 1.
CC Isomerase; Rotamase; Antigen; Signal.
TT SIGNAL
21 POTENTIAL.
```

```

FT CHAIN 22 273 CELL BINDING FACTOR 2.
FT DOMAIN 131 228 PPICT-LIKE.
SQ SEQUENCE 273 AA: 30518 MW: 1C014658BDC4E39 CRC64;

Query Match 13.2%; Score 123; DB 1; Length 273;
Best Local Similarity 28.5%; Pred. No. 0.00092;
Matches 39; Conservative 24; Mismatches 44; Indels 30; Gaps 6;

QY 42 DKEVLNAVIAFKFNNGYPLVNEGQVRYSHLLKNNOSRKPKSMKSPDGISRTDESIQ 101
DB 113 DAARKAFQADQKKDKYVPR-----ARVQAKHLIVATEKEAK-----DIIN 152
QY 102 LKKHLHERLRS---EYKLSLANPES-DCSSHDGCGDGFPSKQGMPPPEEAPFNHY 157
DB 153 ELKG-----LKKELDAFSELAKEKSIDPGSKNGGELGWPDOSTWKPFDAAFAALKN 207
QY 158 GEVSNIIETNSGVHIL 173
DB 208 GRTTTPVKTNFGYHVI 224

RESULT 6
Y175_HELPJ STANDARD; PRT; 299 AA.
ID Y175_HELPJ
AC Q9ZM07;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN JHP0161 PRECURSOR.
GN JHP0161.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OX Helicobacter.
XX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tumming P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -1- SIMILARITY: BELONGS TO THE PPICT/PARYULIN FAMILY OF ROTAMASES.
CC STRONG. TO C.JEJUNI CBE2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
CC EMBL: AE001454; AAD05744.1; -.
DR HSSP: Q13526; IPIN.
DR InterPro: IPR000297; -.
DR Pfam: PF00639; Rotamase; 1.
DR PROSITE: PS01096; PPICT_PPISE.1; 1.
KW Hypothetical protein; Isomerase; Rotamase; Signal.
FT SIGNAL 1 21
FT CHAIN 22 299 HYPOTHETICAL PROTEIN JHP0161.
FT DOMAIN 154 253 PPICT-LIKE.
SQ SEQUENCE 299 AA: 34040 MW: 9C037B1CD1110143 CRC64;

Query Match 13.0%; Score 121.5; DB 1; Length 299;
Best Local Similarity 29.2%; Pred. No. 0.0014;
Matches 40; Conservative 23; Mismatches 45; Indels 29; Gaps 6;

```

```

OY 42 DKEVLNAYIAKFKNNGYKPLVNEGDQVRSVSHLLIKNNOSRKPKSKSPDGSISPTDSEIQ 101
DB 137 EKEMODFYNA-----NKDQLFVKOEAHAR--HILVK-----TEDPAKR 172
OY 102 ILKK-HLERIISGEVKLESLANTES---DCSSHDRCGLGFESKGMOPPEEAAFNLAHV 157
DB 173 IISEIDKOPKAKKEKFEIELAROTIDPNSKNAONGDLGKFGKNOMAPDFSAAFAALTIP 232
OY 158 GEVSNI-IETNSGVHIL 173
DB 233 GDYTKTPVKTERTGYHII 249

RESULT 7
Y175_HELPY STANDARD: PRT: 299 AA.
ID P56112:
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN HP0175 PRECURSOR.
GN HP0175.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
CC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kersavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Richardson D., Dodson R., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson L.M., Lee N., Adams M.D., Hickey E.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.S., Weidman J.M., Fujii C., Bowman C., Matthey L., Mallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -1- SIMILARITY: BELONGS TO THE PPIA/PARVULIN FAMILY OF ROTAMASES.
CC -1- STRONG, TO C.JEJUNI CBE2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000538; AD07245.1;
DR TIGR: HP0175;
DR InterPro: IPR000297;
DR Pfam: PF00639; Rotamase; 1.
DR PROSITE: PS01096; PPIA_PPIASE_1; 1.
DR PROSITE: PS01098; PPIA_PPIASE_2; 1.
KW Hypothetical protein; Isomerase; Rotamase; Signal.
FT SIGNAL 1 21
FT CHAIN 22 299 HYPOTHETICAL PROTEIN HP0175.
FT DOMAIN 154 253 PPIA-LIKE.
FT SEQUENCE 299 AA: 34031 MM; E65F3F2F94B11F5A CRC64;
SQ
Query Match 13.0%; Score 121.5; DB 1; Length 299;
Best Local Similarity 29.2%; Pred. No 0.0014;
Matches 40; Conservative 23; Mismatches 45; Indels 29; Gaps 6;
OY 42 DKEVLNAYIAKFKNNGYKPLVNEGDQVRSVSHLLIKNNOSRKPKSKSPDGSISPTDSEIQ 101
DB 137 EKEMODFYNA-----NKDQLFVKOEAHAR--HILVK-----TEDPAKR 172

```

```

OY 102 ILKK-HLERIISGEVKLESLANTES---DCSSHDRCGLGFESKGMOPPEEAAFNLAHV 157
DB 173 IISEIDKOPKAKKEKFEIELAROTIDPNSKNAONGDLGKFGKNOMAPDFSAAFAALTIP 232
OY 158 GEVSNI-IETNSGVHIL 173
DB 233 GDYTKTPVKTERTGYHII 249

RESULT 8
PPIA_ECOLI STANDARD: PRT: 92 AA.
ID P39159:
AC 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE C (EC 5.2.1.8) (PPIASE C)
DE (ROTAMASE C) (PARVULIN).
GN PPIA OR PARVA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE.
RC MEDLINE=95010704; PubMed=7925971;
RA Rahfeld J.-U., Ruecknagel K.P., Scheibert B., Ludwig B., Hacker J.,
RA Mann K., Fischer G.;
RT "Confirmation of the existence of a third family among
RT recombinant production of parvulin."
RT FEBS Lett. 352:180-184(1994).
RN [2]
RP PRELIMINARY SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
MEDLINE=92358234; PubMed=1379743;
RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT Analysis of the Escherichia coli genome: DNA sequence of the region
RT from 84.5 to 86.5 minutes."
RT Science 257:771-778(1992).
RN [3]
RP REVISIONS, AND IDENTIFICATION.
RC MEDLINE=95184297; PubMed=7878732;
RA Rudd K.E., Sofia H.J., Koonin E.V., Plunkett G. III, Lazar S.,
RA Rouviere P.E.;
RT "Conserved sequence motifs in bacterial and bacteriophage
RT chaperonins."
RL Trends Biochem. Sci. 20:14-15(1995).
RN [4]
RP SEQUENCE OF 1-21, AND CHARACTERIZATION.
RC MEDLINE=94215709; PubMed=8163020;
RA Rahfeld J.-U., Scherhorn A., Mann K., Fischer G.;
RT "A novel peptidyl-prolyl cis/trans isomerase from Escherichia coli."
RT FEBS Lett. 343:65-69(1994).
RN [5]
RP FUNCTION: PPIASE ACCELERATE THE FOLDING OF PROTEINS. IT PREFERENCES
RN AMINO ACID RESIDUES WITH HYDROPHOBIC SIDE CHAINS LIKE LEUCINE AND
RN CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
RN PEPTIDE BONDS IN OLIGOPEPTIDES.
RN -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
RN -1- SIMILARITY: BELONGS TO THE PPIA/PARVULIN FAMILY OF ROTAMASES.
RN -----
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN or send an email to license@isb-sib.ch).
RN -----
DR EMBL: S73874; AA832054.1;
DR EMBL: M67049; AAA67578.1;

```

DR EMBL; AE000454; AAC76780.1; -.  
DR PIR; S45525; S45525.  
DR EcoGene; EG12352; pPIC.  
DR InterPro; IPR000297; -.  
DR Pfam; PF00639; Rotamase; 1.  
DR PROSITE; PS01096; PPIC\_PPIASE\_1; 1.  
DR PROSITE; PS0198; PPIC\_PPIASE\_2; 1.  
KW Isomerase; Rotamase.  
FT INIT\_MET 0  
SQ SEQUENCE 92 AA; 10101 MW; AE1AE202827DF4F CRC64;

Query Match 13.0%; Score 121; DB 1; Length 92;  
Best Local Similarity 37.3%; Pred. No. 0.00033;  
Matches 31; Conservative 13; Mismatches 31; Indels 8; Gaps 2;

OY 99 SIOLIKH-----LELLSGEVLKSLANTESDSCSHDGGDLCFPGSKGMPPFEFA 151  
DB 6 ALHLVKEEKLALDLLEIDIKNG-ADFGKLAKKHSICPSGKRGDLCFPGGQMPVAFDKV 64  
OY 152 AFNLHGEVSNLIETNSGVHLQ 174  
DB 65 VESCVLEPTGTLHTQFGYHIK 87

RESULT 9  
PINL\_HUMAN STANDARD; PRT; 100 AA.  
AC 015428;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PIN1-LIKE PROTEIN.  
GN PIN1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97446133; PubMed=9299231;  
RA Campbell H.D., Webb G.C., Fountain S., Young I.G.;  
RT "The human PIN1 peptidyl-prolyl cis/trans isomerase gene maps to human  
RL chromosome 19p13 and the closely related PIN1 gene to 1p31.";  
RL Genomics 44:157-162(1997).  
CC -1- FUNCTION: NOT KNOWN. COULD BE A TRANSCRIBED PSEUDOGENE.  
CC -1- SIMILARITY: STRONG. TO HUMAN PIN1.  
CC -1- SIMILARITY: CONTAINS 1 MW DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U82382; AAB81333.1; -.  
CC HSSP; O13526; 1PIN.  
DR MIM; 602051; -; PIN.  
DR InterPro; IPR001202; -;  
DR InterPro; IPR002349; -;  
DR Pfam; PF00397; MW; 1.  
DR PRINTS; PRO0403; MWDOMAIN.  
DR PROSITE; PS01159; MW\_DOMAIN\_1; 1.  
FT DOMAIN 5 38 MW.  
SQ SEQUENCE 100 AA; 11021 MW; SDE04C40C0659A56 CRC64;

Query Match 12.2%; Score 114; DB 1; Length 100;  
Best Local Similarity 36.2%; Pred. No. 0.0015;  
Matches 25; Conservative 8; Mismatches 24; Indels 12; Gaps 1;

OY 8 LPPMWTIRSVSRHKEYELNOSTNESSWDPPYGTDEVLNMYIAKFNKNGYKPLVNEGQ 67  
DB 7 LPPGWEKRSRPSRGYGFNHNITNPSQWERPSGN-----SSSGCKIMQGEPAR 54  
OY 68 VRVSHLLIK 76  
DB 55 VRSHLLVK 63

RESULT 10  
PPID\_HAEN STANDARD; PRT; 622 AA.  
ID PPID\_HAEN  
AC P44092;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE D (EC 5.2.1.8) (PPIASE D)  
DE (ROTAMASE D).  
GN PPID OR H11004.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=RD / KM20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shetty R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT influenzae Rd.";  
RL Science 269:496-512(1995).  
RN [2]  
RP CONCEPTUAL TRANSLATION.  
RA Bairoch A.;  
RT Unpublished observations (May-1998).  
CC -1- FUNCTION: PPIASES ACCELERATE THE FOLDING OF PROTEINS. SEEMS TO BE  
CC INVOLVED IN THE FOLDING OF OUTER MEMBRANE PROTEINS (BY  
CC SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC  
CC PEPTIDE BONDS IN OLIGOPEPTIDES.  
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE PPIC/PARVULIN FAMILY OF ROTAMASES.  
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS  
CC INTRODUCED IN POSITION 43 TO PRODUCE THIS ORF.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U32781; AAC22655.1; ALT\_FRAME.  
DR TIGR; H11004; -;  
DR InterPro; IPR000297; -;  
DR Pfam; PF00639; Rotamase; 1.  
DR PROSITE; PS01096; PPIC\_PPIASE\_1; 1.  
DR PROSITE; PS0198; PPIC\_PPIASE\_2; 1.  
KW Isomerase; Rotamase; Transmembrane; Cytoplasmic (POTENTIAL).  
FT TRANSMEM 1 16  
FT DOMAIN 17 37 POTENTIAL.  
FT DOMAIN 38 622 PERIPLASMIC (POTENTIAL).  
FT DOMAIN 270 356 PPIC-LIKE.



SEQUENCE 622 AA: 69590 MW: EE5900A5381C4EB4 CRC64:

Query Match 11.9%; Score 110.5; DB 1; Length 622;  
Best Local Similarity 25.7%; Pred. No. 0.033;  
Matches 39; Conservative 20; Mismatches 54; Indels 39; Gaps 4;

61 LVNEDGVAVSHLLIKNOSRRKPSKSP-----DGISRT-----R 96  
DB 201 LADEMAKOSVDEIKTYEANKSFQPEQYKVOYIDLSDNISRNIQVDAIEAYQ 260  
DY 97 DESIOLKHLERI-----LSGEKLSLANTES-DCSSHDRCGLGFFSK 141  
DY 261 DNKQFOTQHLAHIOFANEODAKVAEELQKGANFADVAKAKSLDKISGENGDGLGWNE 320  
DY 142 GOMQPFEEAFNLHGEVSNIIETNSGVHIL 173  
DB 321 NELPKAFEDAAALGVGYSGPIWVGNYHIV 352

RESULT 11  
NIFM\_AZOV1 STANDARD; PRT: 292 AA.

AC P14890:  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE NIFM PROTEIN.  
GN Azotobacter vinelandii.  
OC Bacteria; Proteobacteria; gamma subdivision; pseudomonadaceae;  
OC Azotobacter.  
OC NCBI\_TaxID=354;

RM SEQUENCE FROM N.A.  
RP MEDLINE-69123097; PubMed-2644218;  
RA Jacobson M.R., Brigle K.E., Bennett L.T., Setterquist R.A.,  
RA Wilson M.S., Cash V.L., Beynon J., Newton W.E., Dean D.R.;  
RT "Physical and genetic map of the major nif gene cluster from  
RT Azotobacter vinelandii."  
RL J. Bacteriol. 171:1017-1027(1989).  
CC -1- FUNCTION: REQUIRED FOR THE ACTIVATION AND STABILIZATION OF THE  
CC IRON-COMPONENT (NIFH) OF NITROGENASE. PROBABLE PIIASE.  
CC -1- SIMILARITY: BELONGS TO THE PIIIC/PARVULIN FAMILY OF ROTAMASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: M20568; AAA64732.1; -  
DR PIR: F32055; F32055.  
DR InterPro: IPR000297; -  
DR Pfam: PF00639; Rotamase; 1.  
DR PROSITE: PS01096; PPII\_PPIASE\_1; 1.  
DR PROSITE: PS01098; PPII\_PPIASE\_2; 1.  
DR Nitrogen fixation; Isomerase; PPII-LIKE.  
KW DOMAIN 148 243  
FT SEQUENCE 292 AA; 32802 MW; 42ED51B1103699BF CRC64;

Query Match 11.3%; Score 105; DB 1; Length 292;  
Best Local Similarity 28.2%; Pred. No. 0.037;  
Matches 29; Conservative 24; Mismatches 40; Indels 10; Gaps 3;

DY 71 SHLLIKNOSRRKPSKSPDGISRRDESIOILKHLERISGEVLSLANTESDCSSH 130  
DY 147 AHLVLTINED-PPENTRE---AARTRIEI-----LKRLGKPERFAQMAKHSECPA 196  
DY 131 DRGDLGFFSKGOMQPFEEAFNLHGEVSNIIETNSGVHIL 173

DB 197 MCGLLGEVVPGLTYPELDACLFQWARGELSPVLESPIGFHVL 239

RESULT 12  
NIFM\_AZOV1 STANDARD; PRT: 293 AA.

AC P23119:  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE NIFM PROTEIN.  
GN Azotobacter chroococcum mcd 1.  
OC Bacteria; Proteobacteria; gamma subdivision; pseudomonadaceae;  
OC Azotobacter.  
OC NCBI\_TaxID=355;

RM SEQUENCE FROM N.A.  
RP MEDLINE-9158323; PubMed-1885524;  
RA Evans D.J., Jones R., Woodley P.R., Wilborn J.R., Robson R.L.;  
RA "Nucleotide sequence and genetic analysis of the Azotobacter  
RT chroococcum nifUSVWZM gene cluster, including a new gene (nifP) which  
RT encodes a serine acetyltransferase."  
RL J. Bacteriol. 173:5457-5469(1991).  
CC -1- FUNCTION: REQUIRED FOR THE ACTIVATION AND STABILIZATION OF THE  
CC IRON-COMPONENT (NIFH) OF NITROGENASE. PROBABLE PIIASE.  
CC -1- SIMILARITY: BELONGS TO THE PIIIC/PARVULIN FAMILY OF ROTAMASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: M60090; AAA22166.1; -  
DR PIR: HA3706; HA3706.  
DR InterPro: IPR000297; -  
DR Pfam: PF00639; Rotamase; 1.  
DR PROSITE: PS01096; PPII\_PPIASE\_1; 1.  
DR PROSITE: PS01098; PPII\_PPIASE\_2; 1.  
DR Nitrogen fixation; Isomerase; Rotamase.  
KW DOMAIN 142 244  
FT SEQUENCE 293 AA; 32978 MW; 7D987E2564DE5CDA CRC64;

Query Match 11.1%; Score 103; DB 1; Length 293;  
Best Local Similarity 26.7%; Pred. No. 0.056;  
Matches 28; Conservative 25; Mismatches 42; Indels 10; Gaps 3;

DY 69 RVSHLLIKNOSRRKPSKSPDGISRRDESIOILKHLERISGEVLSLANTESDCS 128  
DY 146 KARHLVLTINED-PPENTRE---AARTRIEI-----LKRLGKPERFAQMAKHSECP 195  
DY 129 SHDRGDLGFFSKGOMQPFEEAFNLHGEVSNIIETNSGVHIL 173  
DY 196 TAMOGLLGEVVPGLTYPELDACLFQWARGELSPVLESPIGFHVL 240

RESULT 13  
YACD\_BACSU STANDARD; PRT: 297 AA.

AC P37566:  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE HYPOTHEICAL 34.1 KDA PROTEIN IN FTSH-CTSK INTERGENIC REGION.  
GN YACD.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/staphylococcus group; Bacillus.

```

OX NCBL_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
CC -1- SIMILARITY: BELONGS TO THE PPIC/PARVULIN FAMILY OF ROTAMASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC
CC EMBL: D26185; BAA05307.1; -
DR EMBL: Z99104; CAB1848.1; -
DR Subtilist; BG10135; yacd.
DR InterPro: IPR000297; -.
DR Pfam: PF00639; Rotamase; 1.
DR PROSITE: PS01096; PPIC_PPIASE_1; 1.
DR PROSITE: PS50198; PPIC_PPIASE_2; 1.
DR Hypothetical protein; Isomerase; Rotamase.
FT DOMAIN 154 247 PPIC-LIKE.
SQ SEQUENCE 297 AA; 34096 MW; 551B5DEA56457B CRC64;

Query Match 10.5%; Score 97.5; DB 1; Length 297;
Best Local Similarity 21.6%; Pred. No. 0.17;
Matches 40; Conservative 36; Mismatches 62; Indels 47; Gaps 7;

QY 14 IRVSSH-NKEFLNOSTESSMDPPYGTDE-----VLNAVIAK 52
DB 83 LKISSEVDRELLTKAVANSFYEDHTEKWKQDIRNILLEDLTRDIDISNELLS 142
QY 53 FNNNGYKPLVNEGQVRSVSHLLIKNNQSRKPKSMKSPDISRTDESIOILKKHLERILS 112
DB 143 FYNNK-KELYQFDSDSRHNYVKD-----EEARVLYE---LNG 179
QY 113 GEVKLSLANTESDSCSHDGGDLGFFSKQGMPP--FEAAFNLVHGVN-NIETNSG 169
DB 180 GGSFEAVAEERSTDRTSPYGGDLGFVEASDNIPSAYIEAKTLKEDWSEPIKVSNG 239
QY 170 VHILO 174
DB 240 VAIIO 244

RESULT 14
BAG3_MOUSE STANDARD; PRT; 577 AA.
AC 09JULV1; Q9JUC7;
DT 01-OCT-2000 (Rel. 40; Created)
DT 01-OCT-2000 (Rel. 40; Last sequence update)
DT 01-OCT-2000 (Rel. 40; Last annotation update)
DE BAG-FAMILY MOLECULAR CHAPERONE REGULATOR-3 (BCL-2 BINDING ATHANOGEME-
3) (BAG-3) (BCL-2-BINDING PROTEIN BIS).
GN BAG3 OR BIS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBL_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20065084; PubMed=10597216;
RA Lee J.H., Takahashi T., Yasuhara N., Inazawa J., Kamada S.,
RA Tsujimoto Y.;
RT "Bis, a Bcl-2-binding protein that synergizes with Bcl-2 in preventing
cell death.";

```

```

RL Oncogene 18:6183-6190(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Osada N., Kusuda J., Tanuma R., Ito A., Hirata M., Sugano S.,
RA Hashimoto K.;
RT "Isolation of full-length cDNA clones from mouse brain cDNA library
made by oligo-capping method.";
RL Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: INHIBITS THE CHAPERONE ACTIVITY OF HSP70/HSC70 BY
PROMOTING SUBSTRATE RELEASE. HAS ANTI-APOPTOTIC ACTIVITY.
CC -1- SUBUNIT: BINDS TO THE ATPASE DOMAIN OF HSP70/HSC CHAPERONES.
CC BINDS TO BCL-2 AND PLC-GAMMA (BY SIMILARITY).
CC -1- PTM: PHOSPHORYLATED (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 WW DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 BAG DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC
CC EMBL: AF130471; BAA26840.1; -
DR EMBL: AB041583; BAA95066.1; -
DR InterPro: IPR001202; -.
DR InterPro: IPR003103; -.
DR Pfam: PF02179; BAG; 1.
DR PROSITE: PS01159; WW_DOMAIN_1; 1.
DR PROSITE: PS50020; WW_DOMAIN_2; 1.
DR Chaperone; Apoptosis; Repeat; Phosphorylation.
FT DOMAIN 22 56 WW 1.
FT DOMAIN 126 157 WW 2.
FT DOMAIN 426 503 BAG 1.
FT DOMAIN 528 577 BAG 2.
FT DOMAIN 186 193 POLY-SER.
FT CONFLICT 74 74 N -> D (IN REF. 2).
FT CONFLICT 527 527 P -> Q (IN REF. 2).
FT CONFLICT 539 539 K -> E (IN REF. 2).
SQ SEQUENCE 577 AA; 61828 MW; 367528AC914287E1 CRC64;

Query Match 10.4%; Score 97; DB 1; Length 577;
Best Local Similarity 22.9%; Pred. No. 0.45;
Matches 39; Conservative 29; Mismatches 60; Indels 42; Gaps 8;

QY 2 ASTSTGLPPNWTIRVSRSHNKEFLNOSTNESSMD---PYGTDEKVLNAVIAFKNG 57
DB 18 ASDRDLPPGWEIKIDPOTGMFPFVDHNSRTTWDPRVPPGPKDTASSANGP--SRNG 75
QY 58 YKPLVNEG-----QVRVSHL-----LKNNQSRKPKSMK--SPDISRTDESIOILKKH 106
DB 76 SRLPLIREGHPYIPOLRPQYIPVPLVHESSENKRPPLHFAVSPGVQRRRTAAATAPOR 135
QY 107 LERILSGEVLKSLANTESDSCSHDGGDLGFFSKQGM-----QPP 147
DB 136 SOSPLRG--GMTEAAQOTQKC-----GQMPATATTAAGQP 169

RESULT 15
SURA_ECOLI STANDARD; PRT; 428 AA.
AC P21202; P75630;
DT 01-MAY-1991 (Rel. 18; Created)
DT 01-NOV-1997 (Rel. 35; Last sequence update)
DT 01-OCT-2000 (Rel. 40; Last annotation update)
DE SURVIVAL PROTEIN SURA PRECURSOR (PEPTIDYL-PROLYL CIS-TRANS ISOMERASE
SURA) (EC 5.2.1.8) (PPIASE) (ROTAMASE C).
GN SURA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

```

OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Almitron M., Tormo A., Kolter R.;  
 RL unpublished observations (Jan-1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE=9234977; PubMed=1630901;  
 RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,  
 RA Isono K., Mizouchi K., Nakata A.;  
 RT "Systematic sequencing of the Escherichia coli genome: analysis of  
 RT the 0-2.4 min region.";  
 RL Nucleic Acids Res. 20:3305-3308(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=9742617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE OF 331-428 FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE=89359108; PubMed=2670894;  
 RA Roa B.B., Connolly D.M., Winkler M.E.;  
 RT "Overlap between pdxA and ksgA in the complex pdxA-ksgA-apag-apah  
 RT operon of Escherichia coli K-12.";  
 RL J. Bacteriol. 171:4767-4777(1989).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Abe S., Okutsu T., Negishi T., Nakajima H., Aono R.;  
 RT "n-Hexane sensitivity of Escherichia coli due to low expression of  
 RT otaA/omp by insertion of IS2 and identification of the gene product  
 RT as an outer membrane protein.";  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE OF 21-32.  
 RC STRAIN-K12 / EMG2;  
 RX MEDLINE=97443975; PubMed=9298646;  
 RA Link A.J., Robinson K., Church G.M.;  
 RT "Comparing the predicted and observed properties of proteins encoded  
 RT in the genome of Escherichia coli K-12.";  
 RL Electrophoresis 18:1259-1313(1997).  
 RN [7]  
 RP FUNCTION.  
 RX MEDLINE=90330537; PubMed=2165476;  
 RA Tormo A., Almitron M., Kolter R.;  
 RT "surA, an Escherichia coli gene essential for survival in stationary  
 RT phase.";  
 RL J. Bacteriol. 172:4339-4347(1990).  
 RN [8]  
 RP CHARACTERIZATION.  
 RX MEDLINE=96198184; PubMed=8626309;  
 RA Lazar S.W., Kolter R.;  
 RT "SurA assists the folding of Escherichia coli outer membrane  
 RT proteins.";  
 RL J. Bacteriol. 178:1770-1773(1996).  
 RN [9]  
 RP CHARACTERIZATION.  
 RX MEDLINE=97032152; PubMed=8878048;  
 RA Mistakakis D., Belton J.W., Raina S.;  
 RT "New components of protein folding in extracytoplasmic compartments  
 RT of Escherichia coli SurA, FkpA and Skp/OmpH.";  
 RL Mol. Microbiol. 21:871-884(1996).  
 CC -1- FUNCTION: ASSIST IN THE FOLDING OF EXTRACYTOPLASMIC PROTEINS.  
 CC ESSENTIAL FOR THE SURVIVAL OF E. COLI IN STATIONARY PHASE.  
 CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC  
 CC PEPTIDE BONDS IN OLIGOPEPTIDES.

CC -1- SUBCELLULAR LOCATION: PERIPLASMIC.  
 CC -1- SIMILARITY: BELONGS TO THE PPIC/PARVULIN FAMILY OF ROTAMASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.slb.ch](mailto:license@isb.slb.ch)).  
 CC -----  
 CC EMBL: M68521; AAA24304.1; -;  
 CC DR EMBL: D10483; BAA01329.1; -;  
 CC DR EMBL: AE000115; AAC73164.1; -;  
 CC DR EMBL: AB013134; BAA34131.1; -;  
 CC PIR: PV0009; PV0009.  
 CC DR Ecogene: EG10985; SURA.  
 CC DR InterPro: IPR000297; -;  
 CC DR Pfam: PF00639; Rotamase; 2.  
 CC DR PROSITE: PS01096; PPIC\_PP1ASE.1; 2.  
 CC DR PROSITE: PS01098; PPIC\_PP1ASE.2; 2.  
 CC KW Isomerase; Rotamase; Periplasmic; Repeat; Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 1 428 SURVIVAL PROTEIN SURA.  
 FT DOMAIN 171 272 PPIC-LIKE.  
 FT DOMAIN 282 382 PPIC-LIKE.  
 FT CONFLICT 25 25 S -> D (IN REF. 1).  
 FT CONFLICT 116 116 S -> T (IN REF. 1).  
 FT CONFLICT 213 213 G -> GFG (IN REF. 1).  
 FT CONFLICT 315 315 K -> E (IN REF. 2).  
 FT CONFLICT 325 325 A -> T (IN REF. 2).  
 FT CONFLICT 333 333 G -> V (IN REF. 2).  
 SQ SEQUENCE 428 AA; 47283 MW; 25F6ADAB903CDB8E CRC64;

Query Match 9.58; Score 89; DB 1; Length 428;  
 Best Local Similarity 26.48; Pred. No. 1.5;  
 Matches 29; Conservative 23; Mismatches 42; Indels 16; Gaps 4;  
 QY 67 QVRSHTLLNNQSRKRSKSPGDSRTDESIOI-LKKHLERILSGEYKLSPLATES 125  
 DB 284 EVNARHILK-----BSPT--MTDEQARVLEQIADIRKSGKTTFAAAAEFS 329  
 QY 126 -DCSSHDGGLGFFSKQMQPPEEAFLHVGESVNIETNSGVHILQ 174  
 DB 330 QDPESANQGGLGWATPDIDFPAFRDLTRLNKGQMSAPVHSSFGWHLIE 379

Search completed: June 8, 2001, 09:39:14  
 Job time: 79 sec

This Page Blank (uspio)

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2001, 04:01:10 ; Search time 1501.23 Seconds  
(without alignments)  
9715.788 Million cell updates/sec

Title: US-09-507-242-1  
Perfect score: 989  
Sequence: 1 gatcaacacatagatgtgtg.....ggtttgtatattgtgattcc 989

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_ba3: \*  
4: gb\_in1: \*  
5: gb\_in2: \*  
6: gb\_in3: \*  
7: gb\_cm: \*  
8: gb\_ov: \*  
9: gb\_pat1: \*  
10: gb\_pat2: \*  
11: gb\_ph: \*  
12: gb\_pl1: \*  
13: gb\_pl2: \*  
14: gb\_pl3: \*  
15: gb\_pl4: \*  
16: gb\_ba1: \*  
17: gb\_ba2: \*  
18: gb\_fun: \*  
19: gb\_hugo\_hum: \*  
20: gb\_hugo\_inv: \*  
21: gb\_hugo\_rnd: \*  
22: gb\_hug\_hum1: \*  
23: gb\_hug\_hum2: \*  
24: gb\_hug\_hum3: \*  
25: gb\_hug\_hum4: \*  
26: gb\_hug\_hum5: \*  
27: gb\_hug\_hum6: \*  
28: gb\_hug\_hum7: \*  
29: gb\_hug\_hum8: \*  
30: gb\_hug\_hum9: \*  
31: gb\_hug\_hum10: \*  
32: gb\_hug\_hum11: \*  
33: gb\_hug\_hum12: \*  
34: gb\_hug\_hum13: \*  
35: gb\_hug\_hum14: \*  
36: gb\_hug\_hum15: \*  
37: gb\_hug\_hum16: \*  
38: gb\_hug\_hum17: \*  
39: gb\_hug\_hum18: \*  
40: gb\_hug\_hum19: \*  
41: gb\_hug\_hum20: \*  
42: gb\_hug\_hum21: \*  
43: gb\_hug\_hum22: \*

44: gb\_ov: \*  
45: gb\_pat: \*  
46: gb\_ph: \*  
47: gb\_pl: \*  
48: gb\_in: \*  
49: gb\_cm: \*  
50: gb\_ov: \*  
51: gb\_pat: \*  
52: gb\_ph: \*  
53: gb\_pl: \*  
54: gb\_in: \*  
55: gb\_cm: \*  
56: gb\_ov: \*  
57: gb\_pat: \*  
58: gb\_ph: \*  
59: gb\_pl: \*  
60: gb\_in: \*  
61: gb\_cm: \*  
62: gb\_ov: \*  
63: gb\_pat: \*  
64: gb\_ph: \*  
65: gb\_pl: \*  
66: gb\_in: \*  
67: gb\_cm: \*  
68: gb\_ov: \*  
69: gb\_pat: \*  
70: gb\_ph: \*  
71: gb\_pl: \*  
72: gb\_in: \*  
73: gb\_cm: \*  
74: gb\_ov: \*  
75: gb\_pat: \*  
76: gb\_ph: \*  
77: gb\_pl: \*  
78: gb\_in: \*  
79: gb\_cm: \*  
80: gb\_ov: \*  
81: gb\_pat: \*  
82: gb\_ph: \*  
83: gb\_pl: \*  
84: gb\_in: \*  
85: gb\_cm: \*  
86: gb\_ov: \*  
87: gb\_pat: \*  
88: gb\_ph: \*  
89: gb\_pl: \*  
90: gb\_in: \*  
91: gb\_cm: \*  
92: gb\_ov: \*  
93: gb\_pat: \*  
94: gb\_ph: \*  
95: gb\_pl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Pred. No.	Match Length	ID	Description
1	534	54.0	534	13	AF224269
2	99.8	10.1	865	53	AF429018 clone BAO
3	97	9.8	895	12	AF035768 Emericell
4	94.2	9.5	836	6	U78757 Dictyostell
5	81.8	8.3	987	53	AL406159 T3 end of
6	76.8	7.8	951	53	AL394224 T3 end of
7	72.2	7.3	707	14	AJ133755 Digitalis
8	65	6.6	7218	10	166494 Sequence 14
9	64.8	6.6	549	14	AJ06023 Neurospor
10	61.2	6.2	954	94	AB009691 Mus sp. m
11	58	5.9	87947	12	AC006201 Arabidops

12	57.4	5.8	1023	8	AF239760	AF239760 Xenopus 1	
C	13	56.6	5.7	529	14	SCYR017C	
C	14	56.6	5.7	2689	14	SCPEF1	
C	15	56.6	5.7	43661	14	SCXCOSM83	
16	54	5.5	993	6	DMU35140	U35140 Drosophila	
17	53.6	5.4	994	93	HSU49070	U49070 Human pepti	
18	53.6	5.4	1014	9	AR074009	AR074009 Sequence	
19	53.6	5.4	1014	9	AR082120	AR082120 Sequence	
20	52.2	5.3	936	54	CNS07ANM	AL436808 t7 end of	
21	51	5.2	44751	15	SPCC16C4	AL031535 S.pombe c	
C	22	48.8	4.9	38193	65	AC017671	AC017671 Drosophila
C	23	48.8	4.9	298640	4	AE003467	AE003467 Drosophila
24	48.6	4.9	981	2	BS33KDA	X57271 B.subtilis	
25	48.6	4.9	10617	2	BSY14077	Y14077 Bacillus su	
C	26	48.6	4.9	210440	2	BSUB0006	Z59109 Bacillus su
C	27	48	4.9	220808	75	AC018700	AC018700 Homo sapi
C	28	45.8	4.6	186897	63	AC067953	AC067953 Homo sapi
C	29	45.2	4.6	98734	84	PENAL1P2	AL031745 Plasmodiu
C	30	45	4.6	150561	64	AC016457	AC016457 Homo sapi
C	31	45	4.6	229550	2	AP001511	AP001511 Bacillus
C	32	44.8	4.5	59320	4	AC024746	AC024746 Caenorhab
C	33	44.8	4.5	155711	6	YIYD10	AL360354 Plasmodiu
C	34	44.8	4.5	180664	60	AC006706	AC006706 Caenorhab
C	35	44.4	4.5	405	2	AF247686	AF247686 Francisel
36	44.4	4.5	195969	76	AC074354	AC074354 Oryza sat	
37	44.2	4.5	4349	5	AF286897	AF286897 Plasmodiu	
38	44.2	4.5	165213	76	AC074367	AC074367 Homo sapi	
C	39	43.8	4.4	16422	4	AE001421	AE001421 Plasmodiu
40	43.6	4.4	147901	80	AL355587	AL355587 Homo sapi	
41	43.2	4.4	916	53	CNS06090	AL407796 t3 end of	
42	43.2	4.4	169794	60	AC004688	AC004688 Plasmodiu	
43	43	4.3	42043	5	CELC33E10	AF067220 Caenorhab	
C	44	43	4.3	104992	60	AC005504	AC005504 Plasmodiu
C	45	43	4.3	169546	60	AC004157	AC004157 Plasmodiu

## ALIGNMENTS

RESULT	1	AF224269	534 bp	DNA	PLN	02-FEB-2001
LOCUS	AF224269					
DEFINITION						Candida albicans peptidyl prolyl cis/trans isomerase gene, complete cds.
ACCESSION	AF224269					
VERSION	AF224269.1					GI:12655855
KEYWORDS						
SOURCE						
ORGANISM						Candida albicans. Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales; anamorphic Saccharomycetales; Candida.
REFERENCE						
AUTHORS						Devasahayam,G., Chaturvedi,V. and Hanes,S.D.
TITLE						Isolation of the Candida albicans homolog of the ESS1 gene, encoding a prolyl isomerase
JOURNAL						Unpublished
REFERENCE						
AUTHORS						Devasahayam,G., Chaturvedi,V. and Hanes,S.D.
TITLE						Direct Submission
JOURNAL						Submitted (12-JAN-2000) Biomedical Science, Madsenworth
FEATURES						Center/SUNY Albany, 120 New Scotland Avenue, Albany, NY 12208, USA
SOURCE						
location						Qualifiers
1.						534
/organism						"Candida albicans"
/db_xref						"taxon:5476"
<1.						>534
/product						"peptidyl prolyl cis/trans isomerase"
1.						534
/note						"ESS1; Isolated by complementation of Saccharomyces cerevisiae ess1-ts mutant"
/codon_start						=1
/transl_table						=12
/product						"peptidyl prolyl cis/trans isomerase"

Query Match	54.0%:	Score 534:	DB 13:	Length 534:	
Best Local Similarity	100.0%:	Pred. No. 3.3e-99:			
Matches 534:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:	
BASE COUNT	192 a	95 c	114 g	133 t	
ORIGIN					
256	atggcatcgacatacaagcttaccacctaattgacgattagaatcagaatccat	315	/protein_id="AAK00626.1"		
1	ATGGCATCGACATCAACAGCGTTACACCTATTTGACGATTTACATTCAGATCCAT	60	/db_xref="GI:12655856"		
316	aaacaagaattcttctaacaacataccaatgagtcgctcttgagcccaactatg	375	/translation="MASTSTGTPENNTIRVSRSHNKREYFLINSTNESSMDPPYGTDK		
61	AAACAAGAGTATTTCTTAAACCAATCTACCAATGATGCTCTTGGAACCCACTTATG	120	VLNAYIAKFNKNGCYKPLVNDGGVRRVSHLLIKNOSRRKPRSPGSI		
376	actgacaagaagatgattgaaatgcatatcattgacgaagtttaaaacaaatggt	435	LKKHLRRLISGEATKSELANTBEDSCSHRGGDLGFFSKQGMQPPPEEAAPNLHVEV		
121	ACTGACAAGAGATTTGATTCATTCATTCGTAAGTTTAAAAACATGCTTACACCA	180	SNIEFNCSGVHIIQRKG"		
436	cttgatgataagatgagccaggttagagttcttcattctgtatgatacaacataca	495			
181	CTTGATGATGAGATGAGCCAGGTAGAGTTTCATTTGTTGATCAAGAACATCAATCA	240			
496	agaaaacccaagcttctggaagtcctcccgatgattgataagtaagtaagcgaatc	555			
241	AGAAACCCCAAGCTTGTGAAAGTCCCGAGTGTAAAGTAACTAGACAGCAATTCATA	300			
556	cagatattgaaagaacatttgaaagaatattgagttggtggttaacaaatgaaatg	615			
301	CAGATATTGAAGAAACATTGTGAAAGATATTGAGTGTGAGTTAAACATTAAGCAT	360			
616	gcaaatccgaagaattgcatcagctcacatgacagagtggtgatttaggttttagc	675			
361	GCAAAATCCGAAGATGATTCAGCTCAGCTACATGACAGAGTGTGATTAAGGTTT	735			
676	aaagaacaatgcaacacacatctgaaagaagccgcatcaatttgatggttggaagtc	735			
421	AAAGACAAATGACCAACCACTTGAAGAACCCGCAATTCATTTGCAATGTTGAGAA	480			
736	agtaacaatgaaacacatagtggtggtccatatccctccaagaagaagataa	789			
481	AGTAACATTAATTGAACCAATAGTGTGCTCATATCTCTCCAAAGAACAGATAA	534			
RESULT	2				
CNS074N8					
LOCUS	CNS074N8	865 bp	DNA	SNS	11-JAN-2001
DEFINITION					clone BA0AB029808 of library BA0AB from strain CLIB 210 of
ACCESSION	AL429018				
VERSION	AL429018.1				GI:12212212
KEYWORDS					SNS.
SOURCE					Kluveromyces lactis.
ORGANISM					Kluveromyces lactis.
REFERENCE					
AUTHORS					Bolotin-Fukuhara,M., Toffano-Nicche,C., Attiguenave,F.,
TITLE					Duchateau-Nguyen,G., Lemaire,M., Marneise,R., Montocher,R.,
JOURNAL					Robert,C., Termler,M., Wincker,P. and Wesolowski-Louvel,M.
REFERENCE					Genomic Exploration of the Hemiascomycetous Yeasts: 11.
AUTHORS					FEBS Lett. 487 (1), 66-70 (2000)
2					(bases 1 to 865)
Soulet,J.L., Aigle,M., Attiguenave,F., Blandin,G.,					

FEATURES	source
JOURNAL	de-Montigny, J., Dujon, B., Durieux, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuvéglise, C., Ozier, K., Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toftano-Nloche, C., Wesolowski-Louvel, M., Wincker, P., and Weissbach, J.
PUBMED	Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of yeast species for molecular evolution studies(1)
REFERENCE	FEMS Lett. 487 (1), 3-12 (2000)
AUTHORS	11152876
JOURNAL	3 (bases 1 to 865)
COMMENT	Genoscope. Direct Submission Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : segre@genoscope.cns.fr Web : www.genoscope.cns.fr) This STS is part of a random genomic sequencing program of thirteen yeast species: <i>Saccharomyces bayanus</i> var. <i>uvatum</i> , <i>Saccharomyces exiguus</i> , <i>Saccharomyces servazzii</i> , <i>Zygosaccharomyces rouxii</i> , <i>Saccharomyces kluyveri</i> , <i>Kluyveromyces thermotolerans</i> , <i>Kluyveromyces lactis</i> var. <i>lactis</i> , <i>Kluyveromyces marxianus</i> var. <i>marxianus</i> , <i>Pichia angusta</i> , <i>Debaryomyces hansenii</i> var. <i>hansenii</i> , <i>Pichia sorbitophila</i> , <i>Candida tropicalis</i> and <i>Yarrowia lipolytica</i> . Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES	location/Qualifiers
source	1..865
misc_feature	/organism="Kluyveromyces lactis" /strain="CLIB 210" /variety="lactis" /db_xref="taxon:28985" /clone="BA0AB029B08" /clone_id="BA0AB"
misc_feature	<1..>489 /note="similar to <i>Saccharomyces cerevisiae</i> ORF YJR017c [ESL1 : processing/termination factor 1]"
misc_feature	/evidence=not-experimental complement(<12..>226) /note="similar to <i>Saccharomyces cerevisiae</i> ORF YJR018w [questionable ORF]"
misc_feature	/evidence=not-experimental complement(<506..>865) /note="similar to <i>Saccharomyces cerevisiae</i> ORF YGR203w [weak similarity to X.laevi protein-tyrosin-phosphatase cdc homolog 2 and to hypothetical protein YPR200c]"
misc_feature	/evidence=not-experimental
BASE COUNT	265 a 165 c 192 g 243 t
ORIGIN	
Query Match	10.1% Score 99.8; DB 53; Length 865;
Best Local Similarity	60.3%; Pred. No. 9.9e-11;
Matches 202; Conservative	0; Mismatches 127; Indels 6; Gaps 2;
456	ggttaggttcctcatctgtgtgatcaagaacatcaatcaagaacccaagtccttgaa 515
165	GGTCAGAGCCTTGGATCTCTGATTAAAGACACAGAGATCTGGAAGACAGCGTCT--CA 221
516	gtccccagtgtaagtagtagaactagaagaactatcataagatattgaagaagaact 575
222	TAGAAATGAAAAATTTTACCACTTACTTAAGATGAGCGTAAACAGAAATTTAGAACTTTAT 281
576	ggaagaataatctggtgtgtggtctaaactaagtaattgtgcaaatccgaagaagtgtg 635
282	CAAGAG---ATTGAATGTTGGTGAACCACTTCGATTCATTGGCTTAAAGAGCAAGTGA 338
636	caagtcacatgacagaggtgtgtgattttagcagaagaagaatgtcaaacacc 695
339	TTTCATCAGCAAAAGAGGTGATTTGGGGTCTTTCGGTCATCGAAGAAATGCAACCTTC 398
696	attcgaagaagccgcatcaattgtcgtgtgttggaagaagtcagtaacataatgaaccaa 755
399	TTTTCGAAGAGCGTCTTTCCTTGAAGAAATCGATCAAGTCAAGTCAATTTGTAGAAATCTGA 458

Oy	756	tatggtgtccatatcctccaagaaggaata	790
Dd	459	cagtgccttcattcatcaaaagagttccctgaa	493

  

RESULT	3		
LOCUS	AF035768	895 bp	mRNA
DEFINITION	Emeticella nidulans peptidyl-prolyl cis/trans isomerase (plnA)		
ACCESSION	AF035768		
VERSION	AF035768.1	GI:2739196	
KEYWORDS			
SOURCE	Aspergillus nidulans.		
ORGANISM	Aspergillus nidulans.		
REFERENCE	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; Emericella.		
AUTHORS	1 (bases 1 to 895) Crenshaw,D.G., Yang,J., Means,A.R. and Kornbluth,S.		
TITLE	The mitotic peptidyl-prolyl isomerase, Pln1, interacts with Cdc25 and Plxl		
JOURNAL	EMBO J. 17 (5), 1315-1327 (1998)		
MEDLINE	98151356		
REFERENCE	2 (bases 1 to 895) Crenshaw,D.G. and Means,A.R. Direct Submission Submitted (25-NOV-1997) Pharmacology and Cancer Biology, Duke University Medical Center, LIRC Bldg, Lasalle Street Extension, Durham, NC 27708, USA		
JOURNAL	Location/Organisms		
FEATURES	1..895		
source	/organism="Aspergillus nidulans"		
	/strain="R153"		
	/db_xref="taxon:5072"		
gene	1..895		
	/gene="plnA"		
CDS	55..585		
	/gene="plnA"		
	/note="PINA: parvulin type"		
	/codon_start=1		
	/product="peptidyl-prolyl cis/trans isomerase"		
	/protein_id="AAC9984.1"		
	/db_xref="GI:2739197"		
	/translation="MNTVGLPAQMEYVRHSKNLPYPENPATRESRMPEADTMDTLKMYATVHSGAATYHEAPSQBGKIRCSHLVKIKRDSRPSPSEAEITRKKEAREILRGHERIRGEIRLGDLANSEDPSSARKKGIDIGFGRGEMOKFEFEAAFPALPGVYS DIVEGSGRLHIERLO"		
BASE COUNT	252 a	208 c	217 g
ORIGIN	218 t		

  

Query Match	9.8%	Score 97;	DB 12;	Length 895;
Best Local Similarity	51.8%;	Pred. No. 3.7e-10;		
Matches 273;	Conservative 0;	Mismatches 245;	Indels 9;	Gaps 2;

  

Oy	271	acaggctaccaccttaattggacgatagtagtatccagatcccataacaagaatlattc	330
Dd	64	accggtcttccagcagcggttggaagttgccattgcgaactccaAAAAACCCTCCACTAC	123
Oy	331	ttaaaccaattaccaatgatgctgctctggagaccaccttatggcactgtgacaagaagta	390
Dd	124	tttCAACCCCCTTACCAGAGCATCGAGTGGAACCTCCAGCGGATACCGATVTGGAGACT	183
Oy	391	ttgaatgcatacatctggcgaagttaaacaatg-----ttaagaagcacctgtgaaat	444
Dd	184	CTTAAGATGTACATGGCACATATTCACAGCGGCCCGCTACTTACCATGAAAGCCCCGAGT	243
Oy	445	gaggatgtccaggtttaagtttccactattgttgtataagaacaatacaagaanaaaccc	504
Dd	244	CAAGAAGGCAGATCCGTTGCAGTCACCTCCTATTCAAAGCAGACAGACAGAGCGACCG	303
Oy	505	aagctctggaaagtcceccagatggtataagtagaacctagagacgaattctatacagatattg	564

Db 304 AGCAGCTCGAGGAGGAGCGCGAG---ATTACTCGAACAACAAAGAGAGGCGCGCGAGATTTTG 360  
OY 565 aagaacatttggaaagaatatattgagtgtgtgaagttaaacttaagttaattgggaataacc 624  
Db 361 CGCGCTCATCAGAGGCCAATTCATCCGAGGGGAATCCGCGTCGGAGATCTTGGCATGTCC 420  
OY 625 gaaagtattgtagctcaatagacagaggtgtgtgatttaggtttttagaacaagacaa 684  
Db 421 GAGTCGACTCCTCACTAGTGTGCTAGCAAAAGGGCGATCTCGGCTTCTTTGGCCCTGGTGA 480  
OY 685 atgaacaccacattcgaagaagcgcgaatcaatttcattgttggagaagtcgataacata 744  
Db 481 ATGCAGAAAGAGTTCCGAGGAGCGAGCGCTTGCATTGCCAACCGGGTCAAGTCAGTGACATT 540  
OY 745 attaaaccaaatagttgttcgcataatcctcaagaacagatgaat 791  
Db 541 GTCCAAATCAGGGTGTGGTCTTTCATCTTATGAGCGCGCTACAAATAGAT 567

RESULT	4			
DDU78757				
LOCUS	DDU78757	836 bp	mRNA	INV
DEFINITION	Dictyostellium discoideum p1na (p1na) mRNA, partial cds.			
ACCESSION	U78757			
VERSION	U78757.1	GI:1688321		
KEYWORDS	Dictyostellium discoideum.			
SOURCE	Dictyostellium discoideum.			
ORGANISM	Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.			
REFERENCE	1 (bases 1 to 836)			
AUTHORS	Loomis,W.F. and Iranfar,N.			
TITLE	Direct Submission			
JOURNAL	Submitted (19-NOV-1996) Department of Biology, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0302			
	USA			

```

FEATURES
    source
        location/Qualifiers
            1..836
                /organism="Dictyostelium discoideum"
                /strain="AX4"
                /db_xref="taxon:44689"
            <1..733
                /gene="pina"
            <1..733
                /gene="pina"
        /note="Similar to human Pini encoded by the sequence
        presented in Genbank Accession Number U49070"
    CDS

```

```

BASE COUNT      318 a      121 c      127 g      270 t
ORIGIN
/codon_start=2
/product="P1na"
/protein_id="AAB36960.1"
/db_xref="GI:16883322"
/translation="PEWAKPISNVYLEIRKNGVINDRVIDSKDFTVFGRESEVASV
LDHPVSRRHAALVYGANNRFLIDLOSLGTGVNDQYOKPLTPTFVGNFPEKFG
SSKSHFLTKTGTVNPPSSSSSSSSSSSEPKYLCRIILYKHOSRMPSSRSBEKIRTR
TKREAIANKLMDYRRTITISGNTFEDLLHKHNSDCSSAKRGGYLDFKNGQMRPEDECA
FSLKVGVSGLVDDSGVHTIERIA"

```

Query Match 9.5% Score 94.2; DB 6; Length 836;  
 Best Local Similarity 57.4%; Pred. No. 1.4e-09;  
 Matches 190; Conservative 0; Mismatches 130; Indels 3; Gaps 1

QY 468 tcatltgttga tcaagaacaatcactoaagaanaaccccaagcttgtaagtcccaga tgg 527  
| | | | | | | | | | | | | | | | | |  
Db 415 TCATCTTTTAATTAAACATCAGGTTCAAGAAATCCATCTTTGGAG--GAATATAA 471  
| | | | | | | | | | | | | | | | | |  
QY 528 tataagtagaacctgacgcacatctatacagataltgaagaacatttggaaagaatc 587  
| | | | | | | | | | | | | | | | | |  
Db 472 AATTACAAGAACCAGAAAGAAAGCAATCGCTTAACCTTAATGAATATGAGCAACAATTAT 531  
| | | | | | | | | | | | | | | | | |  
QY 588 gagtgtgtgaagtttaactaagtgaatttgccaataaccgaagtgtatgtcacgtacata tga 647

Db	532	ATCTGGTCTCGCACATTTTGAAGTTTAGTCATCAAAAATTCACATTTGTAGCACTGCTAA	591
Qy	648	cagaagtgtagtattgaagtttctttagcaaaagacaatgcaccacatctcgaagaagc	707
Db	592	AAGAGGTGGCATCTGTGATTCATTTCAAAAAGAGGCCAATATGCAAAGACCTTTTGAAGATTG	651
Qy	708	cgatccaattcgatcgtcttgagaagtcagtaacataattgaaaccaatagtgctcca	767
Db	652	CGCATTTCTTTAAAGTTGGTGAAGTAGTAGCGCATTTGTACATCTGATTCAGGTGTTC	711
Qy	768	tatcctccaagaacagcataataccaagata	798
Db	712	TATTTTGAAGATTAGCATTAATAGTATATA	742

RESULT	5
CNS06N09	
LOCUS	CNS06N09 987 bp DNA SFS 10-JAN-2001
DEFINITION	t3 end of clone AU04A008E01 of library AU04A from strain CBS 3080
ACCESSION	AL406159
VERSION	AL406159.1 GI:12169821
KEYWORDS	SFS.
SOURCE	Saccharomyces kluyveri.
ORGANISM	Saccharomyces kluyveri

REFERENCE AUTHORS	TITLE	JOURNAL PUBLISHED
1 (bases 1 to 987)	Neuvéglise, C., Bon, E., Lepingle, A., Wincker, P., Artiguenave, F., Galliard, C. and Casaregola, S.	
	Genomic Exploration of the <i>Hemiascomycetous</i> Yeasts: 9.	
	<i>Saccharomyces kluyveri</i>	
	FEBS Lett. 487 (1), 56-60 (2000)	
2 (bases 1 to 987)		

AUTHORS	TITLE	JOURNAL	REFERENCE
Soucié,J.T., Aigle,M., Artiguenave,F., Blandin,G., Boliotin-Pukhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durans,P., Lepingle,A., Liorette,B., Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S., Sarrin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.	Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of yeast species for molecular evolution studies(1)	FEBS Lett. 487 (1), 3-12 (2000)	3 (bases 1 to 987)

AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Genoscope.  
 Direct Submission  
 Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,  
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
 seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 This STS is part of a random genomic sequencing program of thirteen  
 yeast species: *Saccharomyces bayanus* var. *uvaurum*, *Saccharomyces*  
*exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*,  
*Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces*  
*lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*  
*angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbophilophila*,  
*Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 5 to  
 5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.  
 Location/Qualifiers  
 1. .987

```
misc_feature
<365..+838
/note="similar to Saccharomyces cerevisiae ORF YUR017c.1
ES1; processing/termination factor 1"]
```





```

/codon_start=1
/product="peptidyl-prolyl cis-trans isomerase"
/protein_id="CAB94994.1"
/db_xref="GI:86709932"
/translation="MSSEKVRASHILILKHGCRKSSKDPDGLISATTRDAVSQI
QSLQELSDPASESDILASRSHGSSAKRGDGLGPFGRGQMQKPFEEATFALKGEIS
DIVDTGCVHIKKTG"
BASE COUNT
188 a 185 c 162 g 172 t
ORIGIN

```

Query Match	7.3%	Score 72.2	DB 15	Length 707
Best Local Similarity	66.2%	Pred. No. 4.2e-05		
Matches 104; Conservative	0	Mismatches 33	Indels 0	Gaps 0

07 632 atgacgctcacaatgacagaggtggtatattaaagtgctttttagcacaagacaatgtcaac 6911  
 Db 313 ACTGCAGCTCTGCAAAAGCGTGTGGCGCATCTTGTCCCTTTTGGAAGGGGCCAGATGCAGA 372

**Oy** 692 caccattcgagaagccgcatcattgcagtgtggagaagtcgtaacataatigaa 751  
||| ||||||||| ||| | | ||||| | | ||| |  
**Db** 373 AACCTTTCGAGAAGCCACATTGGCACTAAAGCTTGGTGATTAAGTGATCGTGATA 432

```

52 ccaatagtggtgtccatatatccctccaagaacagagata /88
    | | | | | | | | | | | | | | | | | |
Db 433 CTGACAGTGGAGTTCACATCATCTCAGAGAACACAGGATA 469

```

[illegible]

SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 7218)  
AUTHORS Dörner, F., Schefflinger, F. and Falkner, F. G.ünter  
TITLE Recombinant fowlpox virus  
JOURNAL Patent: US 5670367-A 14-23-SEP-1997;  
FEATURES Location/Qualifiers  
source 1..7218

BASE COUNT	1944 a	1491 c	1486 g	1929 t	368 others
ORIGIN					

Query Match	6.68	Score 65	DB 10	Length 7218
Best Local Similarity	2.18	Pred. No. 0.001		
Matches	8	Conservative	232	Mismatches 137
				Indels 0
				Gaps 0

**OY**    525    tggatataagtagaacctagagacgaattctatacagatatltgaagaaacattgaaaaaat    584  
       |||||    ::::    ::::    :    :    :    :    :    :    :    :    :    :    :    :    :    :  
**Db**    1440    TGGTACRRR    1361

[illegible]

**Oy**    645   tgcacagagtgtgctatttagtggtlttltagcaaaagaacaatgatcacaccatlccaaga     704  
         :: :::: : : : : : : : : : : : : : : : :  
**Db**    1320   RR       1261

```
QY 705 agccgcattcaatttcgacgttgcggaagcagtaacataatgaaccaatcgtgcgt 764
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1260 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1201
```

```
QY 765 ccataccccaagaacagataatacaagatatggaagttgatagaanaaagaanaata 824
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1200 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1141
```

[illegible]

RESULT 9  
NCR6023  
NCR6023  
E40 hp  
mpbn  
PTN  
20-NOV-1998

LOCUS	NC006023	345 bp	MANA	20 NOV 1999
DEFINITION	Neurospora	crassa	mRNA for parvulin (sspl).	
ACCESSION	AJ006023			
VERSION	AJ006023.1	GI:3127914		
KEYWORDS	parvulin; peptidylprolyl isomerase; sspl gene.			
SOURCE	Neurospora	crassa.		
ORGANISM	Neurospora	crassa		

REFERENCE	AUTHORS	TITLE
Sordariales; Sordariaceae; Neurospora.		
1 (bases 1 to 549)		
Tropschug, M.		
Direct Submission		

REFERENCE  
JOURNAL  
Submitted (06-MAY-1998) Tropeschung M., Institut für Biochemie und  
Molekularbiologie, University of Freiburg, Hermann-Herder-Str.  
D-79104 Freiburg, GERMANY  
2 (pages 1 to 549)

TROPSPONG, M.  
 SSPI, A SITE-SPECIFIC PARVULIN HOMOLOG FROM NEUROSPORA CRASSA  
 ACTIVE IN PROTEIN FOLDING  
 J. BIOL. CHEM. 273 (48), 31971-31976 (1998)  
 99041963  
 Location/Qualifiers  
 1..549

```

gene
  /organism="Neurospora crassa"
  /db_xref="taxon:5141"
  /cell_line="74 A"
  1..549
  /gene="sspl"
  1..349
  /gene="sspl"
  /EC_number="5.2.1.8"
  /note="parvulin"

```

/evidence=experimental  
/product="peptidylprolyl isomerase"

```

/db_xref="SPTREMBL:060045"
/translation="MSNTIETGLPEDNEVRHSOKNLPYYENSATKTSRKEPP
DKLTIYMAKYHPTSQOQOQOQOQOQPOGKIRCHALLVKINQSRPPSSWRESEI

```

	KPGESDIVDTASGLHLIERLE"			
BASE COUNT	137 a	174 c	145 g	93 t
ORIGIN				

Query Match	6.6%;	Score 64.8;	DB 14;	Length 549;
Best Local Similarity	48.5%;	Pred. No. 0.0014;		

271 acagcctaccacctaattgacgatatgagfatccagatcccatcaacaagatattc 330

331 ttaaccaatctaccaatgagtcgtctctggaccacacctatgacatgacaagaagta 390

[illegible]



(<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by RepeatMasker (Arjan Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/Mashu/ABI consortium for sequencing BAC clones FBP23, F5J6, T17A5, and T13J16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jining Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khatak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: [at@tigr.org](mailto:at@tigr.org).

FEATURES  
source  
Location/Qualifiers

1. 87947  
/organism="Arabidopsis thaliana"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="II"  
complement(<1. 3627)

misc\_feature  
/note="Sequence from clone T13J16"  
complement(join<343. 462,534. 743,840. 1391,1467. 1745, 1862. 2293,2413. 2841,2921. 3211,3306. 3566,3751. 5049, 5246. 5663,5754. 6166,6574. 6921,7017. 7274,7494. 8144, 8564. 8785,8983. 9204,9310. 9459,9927. 10317,10443. 10665, 10752. 11010,11311. 11646,11740. 12729,13148. 14380, 14459. 14485,14617. 14668,14760. 14823,15211. 15279, 15716. 15775,15858. 15920,16040. 16135,16394. 16502, 16675. 16760,17148. 17863,18061. >18118))  
/gene="At2g17930"  
complement(<343. >18118)

gene  
/gene="At2g17930"  
/note="T13J16.5; similar to hypothetical proteins from S. pombe (C1F5.11C, PID:g1351684), S. cerevisiae (YH9, PID:g731689), and C. elegans (C47D12.1, PID:g1216120", complement(join<343. 462,534. 743,840. 1391,1467. 1745, 1862. 2293,2413. 2841,2921. 3211,3306. 3566,3751. 5049, 5246. 5663,5754. 6166,6574. 6921,7017. 7274,7494. 8144, 8564. 8785,8983. 9204,9310. 9459,9927. 10317,10443. 10665, 10752. 11010,11311. 11646,11740. 12729,13148. 14380, 14459. 14485,14617. 14668,14760. 14823,15211. 15279, 15716. 15775,15858. 15920,16040. 16135,16394. 16502, 16675. 16760,17148. 17863,18061. 18118))  
/gene="At2g17930"  
/note="hypothetical protein"

repeat\_region  
/protein\_start=1  
/protein\_id="A202114.2"  
/db\_xref="GI:6598526"  
/translation="MSPIONFEOHSRLVDLDIPTRLEWVVRDSELIATAEVLT  
NPKAYPPASVILLOITKPOFTDNPBKRNIVELNRLPHSEVLRPPVODLKKA  
MQVLDNENENGLICRTITDLEKNEPTEENQVPLDCKIYSIFRTVSHFEEN  
VKMEVPMKPTSDQSLTPPTPIGWQLNPNSTRSKITTESPLVVMFLFOYSRLV  
QTNIPIILPLAVALISVGPENYPSHLKQVLEKQVLELVSLKHVIGTDFEKRLL  
FLPIDLDERVAVGTGRACFESLRPLAELIYHVVHVGDSLIAOLSLTIFLESN  
MHSTSLISHTTCARLMLNLPPIPEKGVDOOSMDARILLIGRIIDAFAFGKSTPER  
TIIQQLDEGEVGRDRTLRSLDELFPQVPEVHSEKVEDCKNLKTLVMECKTIIMST  
HAHLRPOGKGRHDEVMKASVLSGVHCLALFKKDEKEMLNLFSSIIAIMEPR

misc\_feature

repeat\_region

gene

gene

CDS

DLMDNFSLCPELPESMNNNOUVQFPAALLQAPKRYKRPADVLINLVSSKLDVLKN  
PDSAAKRLVLHFRICIFGAVTKTPDSFELIQHNVINVECKNNTAEKPELGIVQL  
LRVFAAGKCYKXELLRLIPMLPLCMLILMLTGGPGEDKDLLECLTPARQL  
SSLLPYLPRLMKPLVCLSGSDDELVAJLLEFVWDSINPDELPSMANVNSVELA  
LMSHLRPPVPMCKKALQILGKGRNRRKLEPILKECKNDPEHLRLVLPPESTP  
FLYLPDKETINLAAVATQNRHGDYIYRQALKEPLVLCISLOINDEGCYVDGCPRO  
LSTLKSVDSSMNRSEAVEIKADVLKRTQIMARKSTFKTLITLTLAASDPDLS  
TDDEFEVENCIRHAIILHVDYSSNASTSTSSVDGSIYSTSSKSNOSNKQDLP  
IFLDAVLVDYADENRHLAKAALNALVAFETLLFLARVHADVLARGNHSMIVS  
PSNPYSPHSVRIRVPEQLPRLHLGCGSTWQMGVMGALVAGKVVNVEITCY  
FOYKIVRGVLYVLKPLVYASKEEETSOVLMILVNNVNDSEARKRSGVDVE  
YATLEPVPNASTIPKRNQONCALIASRSGSVTEPLTQILLOLIMPLRSEKT  
VDQGVTAALNCLALRPLKVTPELVNFDQAIQABEDTVAAVILMPKYLTS  
LNRLRACIELCTTMAWDFRTQTHNELKATISMEFSLTQRAEIVAAVEGLRQ  
VINQOQMPKELQSSLRPLIVNLAKRNLSMPLLOGLARLLELMSWPNVTLGGKLL  
HLKKWLEPEKLSQKAMKAGEEPKIAAIIELFHLLPHASKEFDELDTLVTLDELA  
LPPGVYSEINSPYRLPLRFLRYAALANDYVLSLSEPKYRRMYLIRSDAGPL  
RELASPKQIISAFPELSPKPDPLSTTAPATSPATSSDDEHISVTKLESSNAVTK  
ANIASDARYFGGLYLITKMYKLIPSWQNSRYVDTLVILKSPARISQNBELNLY  
QVESKFWKVCFLNLYRHEKSEVNVLEDLISTLFSRSDYTFLEKFTIIEVAGETP  
NMRALLLHEPLNLFHSKQGLDHLVQAMQMLIPMLAHFQNGQTEVIDPDKTIV  
ERLIDPEEYSAEYDEPLREILQLATLTKLYQSDLVHHRKELIKFGNNHLKREDA  
SKQMAFVNCHEFDAYOAPERKILLOVFALLFTCGPENKMLVQKALDILMAPKRLP  
LQDSRMPITWRTYKTLIVERGHSIPMLIHFILVLRFLSCRAQFPVPMNSLSR  
LGLPYTTAENRRLAIELAGLYVSRQRQNEKMTDPDGTSLQIDENHTSGADPK  
KSTDSNATSEDEPSKRYKIEPGLDSTICVSGGASSTIPNVEPSATQPODEERKNAM  
EEMIINFLRVALVIEPKDRETMTYKQALDLSQALEWPSANVAFVNLKLLSSMP  
PSQSDSTALQGLDVMKYLKEKOPHLFTRNINQISQILEPCFKHMKIDAGSLSL  
LKWVFAEPFLDAANTPPDKILYOKNELINKNVNTVTAQTDGSDNGSGISFEVLY  
IKTLAVNKNKNDVSVYLRRIORLADISGASHRQROGRTDSAVSTRQATV  
AYICNIRKSVLELIDETVYMLADCKRSYTOILNTLSEKSDASVILCTIDMKRWED  
DPSKTASGLSGSEFTQKQDVLVPLNKLSTIDKQHSESEALEEDQNTYDLDLGLCADS  
TKYPLGLREVSLKVERHEMLGRASHPGRRKRFLLYHELSKTLTARLOYITQID  
WEALSDVFMQLKGLDILLALIVDEDKISLAPNARPLPLSPNPGIOQAQANLEGP  
LEEVTSMEDSIVMKHAOFISATSKLOYADVILPRELHADNANYALMLPEVPIVWT  
LKEBOVALKPMISILSKDYHKQOGRHNVQVLLLEGLOLISHQPRMSELIKIG  
KTYNANHLALLESVYMLPMDNSKCAESIAELYRLNEDNRFGIMKRSTAEIRA  
GDSLVQHGEMQROASLFEYQAMVATQGTNTVYPAKEMCLMEQWMLHCSQSOWAL  
VDRKSIENEVELLIDSLMLKLPMAVATKIDYIPPAQVEPKRLVGSYFALHDNRNG  
VQGAENVTGKGVADLAEQWMOQLEPMESVHARVPLLOOFOQLEVESARLIVDANK  
VSGNTPVGGVGNVADLKQIDLEFMWRLPREPMQNMVWYDMLQMRKMNHYDADKQ  
FATSNRPRIHIGFPRDKAMVNTKLARIARQOKTADVOILKEKYGSTMEVQAFKI  
REOAKYVLEMKGRASGLNINJSTNELYEPDKTAKELFRKQDFHLKLDNBSANAY  
SNAITLTKNLPKMISWGSYCDMAVYQDEEWELEVASCFQDIFRGVSNSSHAR  
VLVLYSFYDANPEVGRVFKHLDQVPHWMLSWIPOLLSTQTEAPHOKLYLKLTA  
VFPOALYMLRTYLLRERDAVNKSELGRVLADROMQONATGAGHGSNLPSENOJOG  
AOTSGAGGTDCGNPHGQESERSTENNLNPGSDQMOHSSSAINNNEVAYNRNAS  
LAISACAPDAADIMBALRGKNNNLASLEVLVLTFRGSEPLVLPRELRVAVNALH  
RCKTPTATTAEVPOPLKEKELSGVCACRSADAVTHAVEVYKODFERHDPDEST  
TFPATYATVLELARLKKNKNILOSNERFPAVLNLEDSRVLRFNVNDEVLPGQYAD  
QEVAPRHYVLELDVGVADPLVVRNDESFRLVLIGSDSQKHFIYQTSLTPNARSDER  
IIEFLFMNMQPHEKESRRRHIGITPIIIPWASVWRMEDMLNFTLEVEYENICA  
RINDRDLRPTTHREKOLNOAISQISAEALRQADYIDTITLNDISFISQVYKT  
LMSGHMMKAKFOANOLAVSSPMSMGIQSGSPKVLPAKTKGMPOTDPRAYDA  
NGMTERNEPFRFLTRMQAFESQFEVGBGLSSKMSAAQAVISSQNIHLRQQLAM  
FROELLSWREPRPLGVPIPEVGGIATLNLPAELKKNANVEDYIKRIADQYFSEE  
DENTVEPPQSVQGVNLEVAALSPNMLCMQPTWHPW"

BASE COUNT	134 a	123 c	164 g	108 t	
ORIGIN	ASORSDSSAAAGDILGSPFGAOKPPEDEASFLALRPEMGSPVFTDGHILLITE				
Query Match	5.8%;	Score 57.4;	DB 8;	Length 529;	
Best Local Similarity	49.6%;	Pred. No. 0.045;			
Matches 176;	Conservative	0;	Mismatches 176;	Indels 3;	Gaps 1.
Oy	445	gagatgcccaggttagatgttctcatttgttgcataagaacaatcaatcaagaataccc	504		
Oy	171	GAGCGGGTAAAGTGCCTGGTGTCTCATCTACTCTTGAGACACAAATCAGCAGGGCCCT	230		
Oy	505	aagcttggaaagtcctccagatgtaataagtagaactagagcaaatctatacagataty	564		
Db	231	TCATCATGTGAGG---CAGACACAAAGTTCACACGCCACCAAGACAGACGCCCTGGA	287		
Oy	565	aagaacaatttggaaaagaatattgagtygttgagglttaactaagtgaattgycaaatcc	624		
Db	288	AATGGCTTCATTCACGAAATGAAATCGGGGGGATGAGGATTTTGAATCTCTTGCTTCAG	347		
Oy	625	gaagatgatttcagctacatgacagcaagtggtgattgaagtttttttagcaagaagaa	684		
Db	348	TTCACTGACTCAGCTCAGCTCAGTCCGAGGGGATCTGGGACCTTTGGCAAGAGTGCA	407		
Oy	685	atgcacaccacatttgcagaagccgcatccaattgtcatlytltggaagatcgatacata	744		
Db	408	ATGCAGAGACCGGTTTGAGATGATCTCTTTCGCTCTGAGAGCCGTGAGAGATGTGGCCC	467		
Oy	745	atgaacacaaatagatggtgtccatatctctccaagaacagagataatacagatat	799		
Db	468	GTCCTTCACAGATCCGGGATTCACATTAATCCTCCGCACTGAGTAATTCACCTTT	522		
RESULT 13					
LOCUS	SCYJR017C	1023 bp	DNA	PLN	11-AUG-1997
DEFINITION	S.cerevisiae chromosome X reading frame ORF YJR017C.				
ACCESSION	Z49517.Y13136				
VERSION	Z49517.1 GI:1015651				
KEYWORDS	baker's yeast.				
SOURCE	Saccharomyces cerevisiae				
ORGANISM	Eukaryote; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
REFERENCE	1 (bases 1 to 1023)				
AUTHORS	de Haan,M., Grievell,L.A. and Smits,P.H.M.				
JOURNAL	Unpublished				
TITLE	2 (bases 1 to 1023)				
FEATURES	MIPS.				
source	Direct Submission				
	Submitted (25-SEP-1995) Data collected by MIPS on behalf of the				
	European yeast chromosome X sequencing project. MIPS at the				
	Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152				
	Martinsried, FRG; E-mail: Mewes@mips.emblnet.org				
	Location/Qualifiers				
	1..1023				
	/organism="Saccharomyces cerevisiae"				
	/db_xref="taxon:4932"				
	/chromosome="X"				
	complement(313..885)				
	/gene="ESS1"				
	complement(313..885)				
	/gene="ESS1"				
	/note="ORF YJR017C"				
	/codon_start=1				
	/protein_id="CAA89541.1"				
	/db_xref="GI:1015652"				
	/translation="MITIVLIPTHNRLRHNDAMPDVA SRGLPTPTWYRSKSKK				
	RKPENPETKHSQEPREGINKDLIHLHDHPRVCAILILIKHRSRPSARSRN				
	ITTSKQDADBLKLITLRDDDSKTNSFEALAKERSDCSSTKRGGDLGWFRGEMOPS				
	FEDAAPLAKGEVSDIVESGSHVIVIRVG"				

CDS 578..940  
/note="ORF YJR018w"  
/codon\_start=1  
/protein\_id="CAA89542.1"  
/db\_xref="GI:1015653"  
/db\_xref="SWISS-PROT:P47091"  
/translation="MSDCLDAGLISLCLMRRCRLHITRPGMSLKLCSLSLVPSGS  
SHCEFPVSGLKRTSLFLDLILYLVHSGSPVLADATSDGISALMCRSLCLVGISTTMI  
IOVLFRLRSKGRYDTRS"

BASE COUNT 259 a 218 c 237 g 309 t

ORIGIN

Query Match 5.7%; Score 56.6; DB 14; Length 1023;  
Best Local Similarity 57.7%; Pred. No. 0.062;  
Matches 101; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 614 tggcaaataccgaagatgattgcagctcacatgacagaagtgtgattgaagttttta 673  
|||||  
Db 488 TGGCTAAGAGAGATCAGATTGCTTTCATACAGCGAGCGGCGACCTCGCTGTTG 429  
gcaagagcaaatgcaaccaccattcgaaagaagcgcattcaattgttgagaag 733  
|||||  
Db 428 GGAAGCGCGAGATCAGCTTACGCTTTGAAGACGCTGCTTCCAGCTCAAGTCCGCGAG 369  
|||||  
Qy 734 tcagtaacataatgaaaccaaattagtggtgtccatatccccaagaacagagata 788  
|||||  
Db 368 TAAGCATATCGTTGATTCAGAGACGCGTGTTCATGTGATCAACGCGGTAGTTA 314

RESULT 14  
SCPTFL 2689 bp DNA 17-FEB-1997  
LOCUS S.cerevisiae Ptf1p gene.  
DEFINITION X85972  
VERSION X85972.1 GI:758283  
KEYWORDS Processing/Termination Factor 1; Ptf1p gene.  
SOURCE Baker's yeast.  
ORGANISM Saccharomyces cerevisiae  
Saccharomycetes; Saccharomycetales;  
Saccharomycetaceae; Saccharomycetes.  
Eukaryota; Fungi; Ascomycota; Saccharomycetes.  
REFERENCE 1 (bases 1 to 2689)  
AUTHORS Engelbrecht,J.A., Voelkel-Melman,K. and Roeder,G.S.  
JOURNAL Meiosis-specific RNA splicing in yeast  
MEDLINE Cell 66 (6), 1257-1268 (1991)  
92005676  
REFERENCE 2 (bases 1 to 2689)  
AUTHORS Hani,J., Stumpf,G. and Domdey,H.  
TITLE Ptf1 encodes an essential protein in Saccharomyces cerevisiae,  
which shows strong homology with a new putative family of p1ases  
JOURNAL FEBS Lett. 365 (2-3), 198-202 (1995)  
95300974  
REFERENCE 3 (bases 1 to 2689)  
AUTHORS Hani,J.  
TITLE Direct Submission  
JOURNAL Submitted (28-MAR-1995) J. Hani, Genzentrum der Universitaet  
Muenchen, Wuermtalstrasse 221, D- 81375 Muenchen, FRG  
4 (bases 1 to 2689)  
REFERENCE 4 (bases 1 to 2689)  
AUTHORS Hani,J.  
TITLE Direct Submission  
JOURNAL Submitted (13-JUN-1995) J. Hani, Genzentrum der Universitaet  
Muenchen, Wuermtalstrasse 221, D- 81375 Muenchen, FRG  
Related sequences: M38340 and P22696.  
COMMENT location/Qualifiers  
FEATURES  
source  
1..2689  
/organism="Saccharomyces cerevisiae"  
/strain="DH484"  
/db\_xref="taxon:4932"  
/clone\_lib="A. Haid"  
/clone="1"  
1..310  
/gene="MER2"  
<1..310

conflict  
/gene="Ptf1p"  
/note="conflict with M38340"  
/citation=[1]  
/replace="c"  
201  
/gene="Ptf1p"  
212..214  
/gene="MER2"  
/note="conflict with M38340"  
/citation=[1]  
/replace="ga"  
554..556  
/note="conflict with M38340"  
/citation=[1]  
/replace="ca"  
584..1633  
/note="unknown ORF"  
/codon\_start=1  
/protein\_id="CAA59960.1"  
/db\_xref="GI:758285"  
/db\_xref="SWISS-PROT:P41903"  
/translation="MSASKMAMNSLEKLELVLPSPTSEVTKYLPAPVGSKGTPEGT  
LVSQSLASLHATVPLNFEPTLSHSEYFKGDPTKITYHQNLRGNFHHKVSAYQ  
HKDLIFTSMTLFAVORSKEHDSLOHETITGLQKOPDPHRYEATSLQKQEVLDPOK  
LSRYASLSDRFODATSMKSYVDAFOYGVMEYOPKDMFYSAHRTDELDYFVKYRPPT  
TVHAGDESSLHKHHPRIKPSITPNDARVNVAVAFYSDSYLLITIPYENLPLYC  
HSFVSIDHTITRYHQLPHVAVNMVITLKISNRSWMDKHLVQGIPTQSGRIASVSQE  
GVVYGSERDIRAKF"  
1809..2587  
/gene="PFP1"  
/gene="PFP1"  
1809..2587  
/gene="PFP1"  
1832..2344  
/gene="Ptf1p"  
1832..2344  
/gene="Ptf1p"  
/codon\_start=1  
/product="Processing/Termination Factor 1"  
/protein\_id="CAA59961.1"  
/db\_xref="GI:758286"  
/db\_xref="SWISS-PROT:P22696"  
/translation="MPSDVASRTGLPPTVTVRYSKSKREYFNPETKHSOWDEPQT  
NKDOLHKLHDPVAVRCHLILIKHSDSRPASHRENTIYSKODATDELKLTITRLD  
DSKTSFELAKERSDSSYKRGDGLWGRGEMQPSFEDAFLQKVEVSDIVSISG  
SGVHVTKRYG"

BASE COUNT 774 a 682 c 596 g 637 t

ORIGIN

Query Match 5.7%; Score 56.6; DB 14; Length 2689;  
Best Local Similarity 57.7%; Pred. No. 0.057;  
Matches 101; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 614 tggcaaataccgaagatgattgcagctcacatgacagaagtgtgattgaagttttta 673  
|||||  
Db 2169 TGGCTAAGAGAGATCAGATTGCTTTCATACAGCGAGCGGCGACCTCGCTGTTG 2228  
gcaagagcaaatgcaaccaccattcgaaagaagcgcattcaattgttgagaag 733  
|||||  
Db 2229 GGAAGCGCGAGATCAGCTTACGCTTTGAAGACGCTGCTTCCAGCTCAAGTCCGCGAG 2288  
|||||  
Qy 734 tcagtaacataatgaaaccaaattagtggtgtccatatccccaagaacagagata 788  
|||||  
Db 2289 TAAGCATATCGTTGATTCAGAGACGCGTGTTCATGTGATCAACGCGGTAGTTA 2343

RESULT 15  
SCXCOSM83 43661 bp DNA PLN 11-NOV-1996  
LOCUS S.cerevisiae chromosome X DNA (cosmid 83).  
ACCESSION X87611  
VERSION X87611.1 GI:854567  
KEYWORDS adenylate cyclase; alpha-agglutinin; ARP sulphurylase;  
beta-adaptin; CN10 region; CoA thioesterase; CYR1 gene;  
dehydroxycid dehydratase; esal gene; glyceroldehyde-3-phosphate  
dehydrogenase; ILV3 gene; metz gene; MET3 gene; sui2 gene; TDH2  
gene.  
SOURCE Baker's yeast.  
ORGANISM Saccharomyces cerevisiae  
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
Saccharomycetaceae; Saccharomycetes.  
REFERENCE 1 (bases 1 to 43661)  
AUTHORS de Haan,M.  
TITLE Direct Submission  
JOURNAL Submitted (25-MAY-1995) M. De Haan, University of Amsterdam, Dept.  
Molecular Biology, Kruislaan 318, NL- 1098 SM Amsterdam,  
NETHERLANDS  
2 (bases 1 to 43661)  
AUTHORS Fang,H., Panzner,S., Mullins,C., Hartmann,E. and Green,N.  
TITLE The homologue of mammalian SPC12 is important for efficient signal  
peptidase activity in Saccharomyces cerevisiae.  
J. Biol. Chem. 271 (28), 16460-16465 (1996)  
96279206  
FEATURES  
SOURCE Location/Qualifiers  
1..43661  
/organism="Saccharomyces cerevisiae"  
/strain="FY1679"  
/sub\_strain="S288C"  
/db\_xref="taxon:4932"  
/chromosome="X"  
/clone="cosmid83"  
1..4065  
/gene="CYR1"  
<1..4065  
/gene="CYR1"  
/codon\_start=1  
/product="adenylate cyclase"  
/protein\_id="CA60917.1"  
/db\_xref="GI:854568"  
/db\_xref="SWISS-PROT:P08678"  
/translation="DPKRHYAIRFENDDFTTLCSTPATVEEIIIPALKRKNITAO  
GNQISLAKVGKSLIKLPTKSPILIEKKLLINKYRKSDPLHMGIEDLSVEYFELH  
PVPSHTPTPEQORIMSEFVHDLRNMDLTPPIIFYQHTSEISLDVSNMNFLEP  
LEFLESSIKLISLMVNIIRASKPSNTTKAYKLSLELRNFIKKVPISIKLSNLT  
LNQCNELESIPAGVELKLNQLDLSSNKFMYPEVINYCTNLQIDLSYKLSQSP  
OSTYLYLAKRMNSHNKLNFIQDLSMTDLRTLNIRRSIKTANSNLCNFID  
NRISNPDPLPKLALIEQENPTISFKOPYRNKMSLTINKQSLISIPBELTKLS  
PLEKELNNQNLRLPQESIKTLVFLSVARNLEVIPELSQLSLRTIDLSNNT  
RDFVQGMENLELTSLSNNAFNSLSNFSYSHKLESITELYSGNKLITLGGTVAKMSLKT  
MMPLEFNCVNLKYNLSYNNFSDVSHKLESITELYSGNKLITLGGTVAKMSLKT  
LMAINSQMLSLPAELSNLSOLSVFVDCANOLKYNISNHYDMNRNKKELKLYNEN  
RRFEIKSFISHDIDLTLVLPOLKVLGMQVTLNMTTKVDPDENVNRLPTASTIN  
GMRYGVDITLQGRYVSSRDYTFERFRGNDECLICLHDSKNQNDGHNISRIYROI  
YDKTLRQLERYGDETNDIKTALRFSFLQINKELNLSVNGADVANSIVDLVS  
GACSTVIYIRGKLFKLANLGDCAILSKNNGDXYOTLKQHLPTKREVEYRIISGVY  
NNGLDGVNPLSRAVGFEDLPHIHASPDISVTLTKADEMLIVATHKLMYMDVY  
CDIARENSDTPLRAAELKDHAAYGCTENITILCALYENIQOONRPTLNKNSLMFR  
LRTEFEVTLRIQPEISPTPCNLAMVFTDIKSTFLELFPNARTRIAKTINDIMRRO  
LRITGTEVETEGDAFWARPTPTSGLWCLSVOLKLDQWPEEITSVQGCQVTR  
NGNITVGLSVRMGIHWGCPVPELDLVTQRMDYLGPNVKNARQVAGVADGQAMSSD  
FYSEFNKIMRYHERVVGKSKLEVEGEELIGEVLEERIMLESIGWAFDFGKHLK  
GLETEKLVITLAYPKILASRHEFASDEDSKLIETMLERLIVISNRLSISMSALSGF  
LEIDSRTEGYSIKRPNKVENGMOSISEKALLFEFDHYITRIESSVALHLHQRCSC  
LEICRNDKTSARSNIENVDELLOMVKNADLSI"  
complement(4417..5028)  
/gene="ORF YJR83.17"  
/gene="ORF YJR83.17"  
complement(4417..5028)  
CDS

gene  
CDS  
gene  
CDS  
gene  
CDS  
gene  
CDS  
variation  
gene  
CDS  
gene  
CDS  
/gene="ORF YJR83.17"  
/codon\_start=1  
/protein\_id="CA60918.1"  
/db\_xref="GI:854569"  
/db\_xref="SWISS-PROT:P41544"  
/translation="MVSIRRLVRPNELKPSQIKQDSLSPSKIGLOIYLIIQFYTT  
AIVLVCWAKLAGYDLNIKEMLEFSMENIDFTNAGLSLSLMLDLSLCVFFLVYIG  
RSKLAMPFAITIIHAINFIVFLYTRKPEFSWFELQILSLILFLITWTRMELMD  
TFEGLVDPNNGEVLVTPSQOHSNHSLEOSP1QLDLDESQI"  
5472..5828  
/gene="ORF YJR83.2"  
5472..5828  
/gene="ORF YJR83.2"  
/codon\_start=1  
/protein\_id="CA60919.1"  
/db\_xref="GI:854570"  
/db\_xref="SWISS-PROT:P47081"  
/translation="MSPSGKKFRSRROOLYVEASLAGRYKALKSHPLFGCLPECAT  
IVLSFWLSSTFAITRYEGDKRVOEINEDILKTRKNOREFDIIEEYTRLOGLSEDM  
EPVRVARLKDESNW"  
complement(6049..7479)  
/gene="RIB1"  
complement(6049..7479)  
/note="homologue of the human /rat ribophorin I precursor  
protein RIB1"  
/codon\_start=1  
/protein\_id="CA60920.1"  
/db\_xref="GI:854571"  
/db\_xref="SWISS-PROT:P41543"  
/translation="MROVFSWIVGLFCEPCFNSSAAOYEPPTAWEVDKRTIDVS  
AYISEETLEITKINLSEPAETEYFAFEGISKVSFSAVFTNEATEPLNSOLANST  
APGDGESEIRYGIIOFNALISPOEVSILKSYFNAGYIPPEHVMSEQOHLMET  
NRPLSADKTKASFITLIGSSFEYHPNDSILGANGSNFEEPMEDIPRSSNE  
TLAIVYSNAPLNOVNNLRDIMSHTASTOFEVEYELTKAKLSGFSRLGMO  
IQTONMGTRHYVTLDMILPEGARDHPTFDYGLVYSNASHARDFTIRPRPITGMO  
INFTVGMTRKLSDFLHVSSGDEKFAVSIPLNCPDVTVDNELSLFPGALTFDI  
DSPVPTNVSLETQKSYFDLNGKHKTFESYRNLSIOVANGQVLIKDYPRKSPFKP  
LSIACYITFALMGVFLKTLNMNVTN"  
8173..8754  
/gene="proteasome component pre3"  
8173..8754  
/gene="proteasome component pre3"  
/codon\_start=1  
/protein\_id="CA60921.1"  
/db\_xref="GI:854572"  
/db\_xref="SWISS-PROT:P38624"  
/translation="MAYTFKQGVILGADSRITTGATYANRYTDKLTVRHDKINCRCBS  
SAADTQATADIVQYHLLEYTSQYGPSTETNASVFKELCYENKDNLTAGIIVAGYDK  
NKGVYITPLGGSVHKLPYAIAGSGSTFIYGVCDKNRENNKSEETVDFTKHLSQAI  
KMDSSGGVIRMVULTAGVERLIFYPDEYQL"  
complement(9162..9223)  
/note="CEN10 (centromeric region)"  
9630..11438  
/gene="ORF YJR83.4"  
9630..11438  
/gene="ORF YJR83.4"  
/codon\_start=1  
/protein\_id="CA60922.1"  
/db\_xref="GI:854573"  
/db\_xref="SWISS-PROT:P47082"  
/translation="MPDEPLSPGMRKREYHYISIPLNRSASFSPDSYSQFOSDF  
MTRROSLIDHVGSGFKGNSLSRATSLIRANSNPTIRLEANDNRSFKESEDNEDTYP  
DTLAPALDGRSLSTLNAGGRPTTNLANDRVSATSMALIHDDYGSIONSTGDS  
ILRPASLTEMSSGAGRRPTNNMDSTIVARVGVGKVTLLDAGSTLADSTBOTIFNSI  
NVLIGIGLALPLKLYAGWYIGTMAIFALAFCAELSLRSLDLDPLIIVADIG  
VYAGGTGRALISLFTLDLIGSVSVIILFGDLSLNLFPQYSTPFKIVSFFLVTPP  
VFIFLVSLSNLSLIGTSTGTIVYICCGGKYSKSSGSLVNPETMMPIDKLCLL  
SIGLSMGWGHAVFPMKTDMRHMDKPKDCLTKTYKITSTVDIGTAVIGIMGNLY  
KDEITKAVLLEGGPKFYGLISLNTIIPAKTIPUNARPIVSVLDVLMVQIHDEA  
SAIKRRAKGLQVFNRIFINVFLAINPEEPKIIAFGAGLCTITCLILPCWFYL  
RLCKTTITKPMERVACHVITICISVYLSLGVGAALIS"  
11692..13473  
gene





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using SW model

```
Run on:      June  8, 2001, 04:38:56 ; Search time 217.33 Seconds
              (without alignments)
              2656.604 Million cell updates/second
```

Title:	US-09-507-242-1
Perfect score:	989
Sequence:	1 gatcaaccatagatgttgt.....ggtttgtatattgatcc 989

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

```
Searched:      678276 segs, 291890651 residues
Total number of hits satisfying chosen parameters: 1356552
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match 0%
	Maximum Match 100%
	Listing first 45 summaries

Database

N\_genesec\_U40v1.1

- 1: /SID56/gcgdata/genesec/geneseq/NA1980.DAT.\*
- 2: /SID56/gcgdata/genesec/geneseq/NA1981.DAT.\*
- 3: /SID56/gcgdata/genesec/geneseq/NA1982.DAT.\*
- 4: /SID56/gcgdata/genesec/geneseq/NA1983.DAT.\*
- 5: /SID56/gcgdata/genesec/geneseq/NA1984.DAT.\*
- 6: /SID56/gcgdata/genesec/geneseq/NA1985.DAT.\*
- 7: /SID56/gcgdata/genesec/geneseq/NA1986.DAT.\*
- 8: /SID56/gcgdata/genesec/geneseq/NA1987.DAT.\*
- 9: /SID56/gcgdata/genesec/geneseq/NA1988.DAT.\*
- 10: /SID56/gcgdata/genesec/geneseq/NA1989.DAT.\*
- 11: /SID56/gcgdata/genesec/geneseq/NA1990.DAT.\*
- 12: /SID56/gcgdata/genesec/geneseq/NA1991.DAT.\*
- 13: /SID56/gcgdata/genesec/geneseq/NA1992.DAT.\*
- 14: /SID56/gcgdata/genesec/geneseq/NA1993.DAT.\*
- 15: /SID56/gcgdata/genesec/geneseq/NA1994.DAT.\*
- 16: /SID56/gcgdata/genesec/geneseq/NA1995.DAT.\*
- 17: /SID56/gcgdata/genesec/geneseq/NA1996.DAT.\*
- 18: /SID56/gcgdata/genesec/geneseq/NA1997.DAT.\*
- 19: /SID56/gcgdata/genesec/geneseq/NA1998.DAT.\*
- 20: /SID56/gcgdata/genesec/geneseq/NA1999.DAT.\*
- 21: /SID56/gcgdata/genesec/geneseq/NA2000.DAT.\*
- 22: /SID56/gcgdata/genesec/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	974.6	98.5	989	21	A50292	Candida albicans C
2	87.2	8.8	747	21	C33551	Arabidopsis thalia
3	87.2	8.8	833	21	C47563	Arabidopsis thalia
4	87.2	8.8	835	21	C36568	Arabidopsis thalia
5	53.6	5.4	1014	18	168888	NIMA-interacting p
6	53.2	5.4	662	21	F07816	Fusarium venenatum
7	52.4	5.3	406	21	F11336	Aspergillus niger
8	50	5.1	7963	18	V74479	Staphylococcus aur
9	46.4	4.7	580	20	Z33510	Human prostate can
10	38.6	3.9	666	21	C35259	Arabidopsis thalia
11	38.4	3.9	910725	20	X20248	Borrelia burgdorferi

12	38.2	3.9	4468	21	C62126
13	38.2	3.9	4611	21	C62127
14	38.2	3.9	1038602	20	Z01425
C 15	37.6	3.8	752	21	C68123
C 16	37.2	3.8	2146	21	C36046
C 17	37.2	3.8	3265	21	Z65095
C 18	37	3.7	20284	21	C69339
C 19	36.6	3.7	11922	21	A70187
C 20	36.2	3.7	296	14	O60633
C 21	36.2	3.7	1204	21	C83102
C 22	36.2	3.7	1449	21	A70251
C 23	36.2	3.7	19124	18	T72882
C 24	36.2	3.7	19124	21	Z89827
C 25	36.2	3.7	19176	19	V21209
C 26	36	3.6	166478	19	
C 27	36	3.6	979	20	X61806
C 28	36	3.6	1047	20	X61805
C 29	36	3.6	2892	20	X20292
C 30	36	3.6	4810	20	V70887
C 31	35.8	3.6	4838	20	V70891
C 32	35.6	3.6	3541	19	V14522
C 33	35.6	3.6	1258	21	C59345
C 34	35.6	3.6	1511	15	O28302
C 35	35.6	3.6	1511	19	O66798
C 36	35.6	3.6	1511	20	Z10082
C 37	35.6	3.6	1511	19	V14508
C 38	35.6	3.6	1511	19	V14508
C 39	35.6	3.6	1511	20	Z10082
C 40	35.4	3.6	1879	21	C75943
C 41	35.4	3.6	3549	21	A70223
C 42	35.4	3.6	1397	22	F22291
C 43	35.4	3.6	15826	21	F23002
C 44	35.4	3.6	7753	15	O56826
C 45	35.2	3.6	27354	21	C81914
		3.6	1230025	20	X91999
		3.6	201	15	O66818
		3.6	4041	21	A70170

## ALIGNMENTS

XX	RESULT	1
XX	A50292	
ID	A50292 standard; DNA; 989 BP.	
AC	A50292;	
XX		
DT	20-NOV-2000 (first entry)	
XX		
DE	Candida albicans CaESL1 gene.	
XX		
KX	CaESL1; ESLS1; Infection; diagnosis; therapy; antifungal; fungicide	
KM	antiproliferative; cytostatic; antitumour; ds.	
XX		
OS	Candida albicans.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	256..789
FT	/tag= a	
XX		
PN	W0200050561-A2.	
XX		
PD	31-AUG-2000.	
PF	18-FEB-2000; 2000MO-US04203.	
XX		
PR	23-FEB-1999; 99US-0121246.	
XX		
PA	(HEAL-) HEALTH RES INC.	
XX		
PI	Hanes SD, Devasahayam G, Chaturvedi V;	
DR	WPI: 2000-565453/52.	
XX	P-PsDB; Y95876.	
XX		

PT Novel Candida albicans gene, CaESS1 useful for identifying compounds  
PT that specifically bind to and/or inhibit CaESS1 and thus for treating  
PT Candida albicans infections and other life-threatening fungal  
PT infections -  
XX  
PS Claim 2; Fig 1A; 51pp; English.  
XX  
CC This is the complete nucleotide sequence of the CaESS1 gene from  
CC Candida albicans. The gene encodes a 177-amino acid protein,  
CC CaESS1 (see Y95876), which is the C. albicans homologue of  
CC Saccharomyces cerevisiae ESS1 (42% amino acid identity). The  
CC CaESS1 gene was isolated from a C. albicans genomic DNA library by  
CC functional complementation of a temperature-sensitive S. cerevisiae  
CC strain, ess1-L94pts. CaESS1 nucleic acids, especially CaESS1-specific  
CC primers (see A50293-94) and probes to determine the presence of C.  
CC albicans in a sample or specimen. CaESS1 protein is a target  
CC for screening for antifungal and inhibitor compounds, useful for  
CC treating or preventing C. albicans infections. CaESS1 DNA can also  
CC be used to generate diagnostic probes or primers for replicating or  
CC cloning C. albicans DNA. A CaESS1 inhibitor can be a compound  
CC which selectively inhibits growth of S. cerevisiae not containing an  
CC endogenous ESS1 gene but rather CaESS1 and uninduced PIN1 (a human  
CC homologue of ESS1) and/or preferably does not inhibit induced PIN1,  
CC e.g. does not inhibit S. cerevisiae not containing an endogenous  
CC ESS1 but rather induced PIN1. Compositions which inhibit PIN1  
CC are useful antiproliferatives e.g. antineoplastics, antitumour  
CC agents or anticancer agents.  
XX  
XX  
SQ Sequence 989 BP; 340 A; 174 C; 190 G; 285 T; 0 other;

Query Match 98.5%; Score 974.6; DB 21; Length 989;  
Best Local Similarity 99.1%; Pred. No. 2.8e-236;  
Matches 980; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 gataaccaataagatgtgtgtgtctaaccaagtaacagcggttgaagacaagagttatt 60  
DB 1 gataaccaataagatgtgtgtgtctaaccaagtaacagcggttgaagacaagagttatt 60  
QY 61 tagacacacagcattttagctacttgatagatatacagtttagcttctctgcaataga 120  
DB 61 tagacacacagcattttagctacttgatagatatacagtttagcttctctcaataga 120  
QY 121 tattaagtagtgcatttaccacaaactctctcttcttatacttatacttcaacaaca 180  
DB 121 tattaagtagtgcatttaccacaaactctctcttcttatacttatacttcaacaaca 180  
QY 181 agatttcgtgtgcttctgtgtatatttgcacagtttagcttctcttctt 240  
DB 181 agatttcgtgtgcttctgtgtatatttgcacagtttagcttctcttctt 240  
QY 241 cagtagtatacatcatgtgcacatcaacagcgttaccacctaattgacgattaga 300  
DB 241 cagtagtatacatcatgtgcacatcaacagcgttaccacctaattgacgattaga 300  
QY 301 gtatccagatccataaacaagagatttcttaaccaaatctacacatgagatctctg 360  
DB 301 gtatccagatccataaacaagagatttcttaaccaaatctacacatgagatctctg 360  
QY 361 gaccacacctatgacacgacaaagagatttgaatcatcatgtgcgaatttaaaac 420  
DB 361 gaccacacctatgacacgacaaagagatttgaatcatcatgtgcgaatttaaaac 420  
QY 421 aatgttcaagcactgtgtaagagatgacaggttagagtttccatttcttctgac 480  
DB 421 aatgttcaagcactgtgtaagagatgacaggttagagtttccatttcttctgac 480  
QY 481 aagaacaatcaataagaanaacccaagctctggaagtcaccagatggtataagtaact 540  
DB 481 aagaacaatcaataagaanaacccaagctctggaagtcaccagatggtataagtaact 540  
QY 541 agagaacgaattatacagattttgaagaacaattggaagaataatgagtggtgagtt 600  
DB 541 agagaacgaattatacagattttgaagaacaattggaagaataatgagtggtgagtt 600

DB 541 agagaacgaattatacagattttgaagaacaattggaagaataatgagtggtgagtt 600  
QY 601 aaactaagtgaattggaacataaccgaaagtgattgcagctcacatgacaagtgatg 660  
DB 601 aaactaagtgaattggaacataaccgaaagtgattgcagctcacatgacaagtgatg 660  
QY 661 ttaaggtttttaaagaagaacaatgcaaccacattcgagaagccgcatcaatttg 720  
DB 661 ttaaggtttttaaagaagaacaatgcaaccacattcgagaagccgcatcaatttg 720  
QY 721 catgttgagaagtcagtaacataattgaaacccaatggtgtgtccatctccaaaga 780  
DB 721 catgttgagaagtcagtaacataattgaaacccaatggtgtgtccatctccaaaga 780  
QY 781 acaagataacaaagataattgagttgattcatataaataatgaagacaagtgtg 840  
DB 781 acaagataacaaagataattgagttgattcatataaataatgaagacaagtgtg 840  
QY 841 atagattgtgaacccaaaagcgatgctcacaagaatcgaaaaactgtgagaagaat 900  
DB 841 atagattgtgaacccaaaagcgatgctcacaagaatcgaaaaactgtgagaagaat 900  
QY 901 ctaccaggtacacggcgattaaacttaactgctgatatattatgaatcgagaagtt 960  
DB 901 ctaccaggtacacggcgattaaacttaactgctgatatattatgaatcgagaagtt 960  
QY 961 ccgcgtacgtgttgttatatttgatcc 989  
DB 961 ccgcgtacgtgttgttatatttgatcc 989

RESULT 2  
C33551  
ID C33551 standard; DNA; 747 BP.  
XX  
AC C33551:  
XX 17-OCN-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 3467.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
XX  
FN EP1033405-A2.  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999;	99US-0132487.
PR 07-MAY-1999;	99US-0132863.
PR 11-MAY-1999;	99US-0134256.
PR 14-MAY-1999;	99US-0134218.
PR 14-MAY-1999;	99US-0134219.
PR 14-MAY-1999;	99US-0134221.
PR 14-MAY-1999;	99US-0134370.
PR 18-MAY-1999;	99US-0134768.
PR 19-MAY-1999;	99US-0134941.
PR 20-MAY-1999;	99US-0135124.
PR 21-MAY-1999;	99US-0135353.
PR 24-MAY-1999;	99US-0135629.
PR 25-MAY-1999;	99US-0136021.
PR 27-MAY-1999;	99US-0136392.
PR 28-MAY-1999;	99US-0136782.
PR 01-JUN-1999;	99US-0137222.
PR 03-JUN-1999;	99US-0137528.
PR 04-JUN-1999;	99US-0137502.
PR 07-JUN-1999;	99US-0137724.
PR 08-JUN-1999;	99US-0138094.
PR 10-JUN-1999;	99US-0138540.
PR 10-JUN-1999;	99US-0138847.
PR 14-JUN-1999;	99US-0139119.
PR 16-JUN-1999;	99US-0139452.
PR 16-JUN-1999;	99US-0139453.
PR 17-JUN-1999;	99US-0139492.
PR 18-JUN-1999;	99US-0139454.
PR 18-JUN-1999;	99US-0139455.
PR 18-JUN-1999;	99US-0139456.
PR 18-JUN-1999;	99US-0139457.
PR 18-JUN-1999;	99US-0139458.
PR 18-JUN-1999;	99US-0139459.
PR 18-JUN-1999;	99US-0139460.
PR 18-JUN-1999;	99US-0139461.
PR 18-JUN-1999;	99US-0139462.
PR 18-JUN-1999;	99US-0139463.
PR 18-JUN-1999;	99US-0139750.
PR 18-JUN-1999;	99US-0139763.
PR 21-JUN-1999;	99US-0139817.
PR 22-JUN-1999;	99US-0139899.
PR 23-JUN-1999;	99US-0140353.
PR 23-JUN-1999;	99US-0140354.
PR 24-JUN-1999;	99US-0140695.
PR 28-JUN-1999;	99US-0140823.
PR 29-JUN-1999;	99US-0140991.
PR 30-JUN-1999;	99US-0141287.
PR 01-JUL-1999;	99US-0141842.
PR 01-JUL-1999;	99US-0142154.
PR 02-JUL-1999;	99US-0142055.
PR 06-JUL-1999;	99US-0142390.
PR 08-JUL-1999;	99US-0142803.
PR 09-JUL-1999;	99US-0142920.
PR 12-JUL-1999;	99US-0142977.
PR 13-JUL-1999;	99US-0143542.
PR 14-JUL-1999;	99US-0143624.
PR 15-JUL-1999;	99US-0144005.
PR 16-JUL-1999;	99US-0144085.
PR 16-JUL-1999;	99US-0144086.
PR 19-JUL-1999;	99US-0144325.
PR 19-JUL-1999;	99US-0144331.
PR 19-JUL-1999;	99US-0144332.
PR 19-JUL-1999;	99US-0144333.
PR 19-JUL-1999;	99US-0144334.
PR 19-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144352.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144884.
PR 21-JUL-1999;	99US-0144814.
PR 21-JUL-1999;	99US-0145086.
PR 21-JUL-1999;	99US-0145087.
PR 22-JUL-1999;	99US-0145085.
PR 22-JUL-1999;	99US-0145087.
PR 22-JUL-1999;	99US-0145089.
PR 22-JUL-1999;	99US-0145192.
PR 22-JUL-1999;	99US-0145145.
PR 23-JUL-1999;	99US-0145218.
PR 23-JUL-1999;	99US-0145224.
PR 26-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 27-JUL-1999;	99US-0145919.
PR 28-JUL-1999;	99US-0146385.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 02-AUG-1999;	99US-0146389.
PR 03-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 04-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 05-AUG-1999;	99US-0147260.
PR 06-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 09-AUG-1999;	99US-0147935.
PR 10-AUG-1999;	99US-0148171.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 13-AUG-1999;	99US-0148684.
PR 16-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149175.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149902.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150566.
PR 26-AUG-1999;	99US-0150884.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 27-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155139.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 06-OCT-1999;	99US-0157865.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0158293.
PR 13-OCT-1999;	99US-0158294.
PR 13-OCT-1999;	99US-0159295.
PR 14-OCT-1999;	99US-0159329.
PR 14-OCT-1999;	99US-0159330.
PR 14-OCT-1999;	99US-0159331.
PR 14-OCT-1999;	99US-0159637.
PR 14-OCT-1999;	99US-0159638.
PR 18-OCT-1999;	99US-0159584.
PR 21-OCT-1999;	99US-0160741.
PR 21-OCT-1999;	99US-0160767.
PR 21-OCT-1999;	99US-0160768.
PR 21-OCT-1999;	99US-0160770.
PR 21-OCT-1999;	99US-0160814.

PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match		
Best Local Similarity		8.8%;
Matches 207; Conservative		53.9%;
		Pred. No. 8.2e+13;
		Mismatches 168;
		Indels 9;
		Gaps 1;

PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.

PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159332.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159588.  
PR 18-OCT-1999; 99US-0159589.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 8.8%; Score 87.2; DB 21; Length 833;  
Best Local Similarity 53.9%; Pred. No. 8.5e-13;  
Matches 207; Conservative 0; Mismatches 168; Indels 9; Gaps 1;

QY 415 aaaaacatggttaacagccattgtgaatgagatgagcaggttaagattcattcattg 474  
DB 143 aacaagcatcgaaagaagagcaagatggtcgcagagaccagttaaagcatcgacatt 202  
QY 475 ttgatcaagaacatcaatcaagaagaaccgaagctcttggaatcccgagatg----- 527  
DB 203 ttgatcaagaacatcaatcaagaagaaccgaagctcttggaagatcccgagatgatt 262  
QY 528 --tatagttagaactagaacgaatctatcacagatatitgaagaacattggaagaata 585  
DB 263 attcgactaccaccatcagaagacgcgcgttgtagcagcttaaatcgatccgtgaagatatt 322  
QY 586 ttgagtggtgaggttaactaagtgaattggaatgcaaataccgaagatgattgcagctcacat 645  
DB 323 gtcctcggaagcggaatttcgaagaagatgctgattctgactgattgactgtcgt 382  
QY 646 gacaagatggtgatttggtgttttttagcaagaagacaatgcaaccacccattggaaga 705  
DB 383 aaacgcgcgcgtgattctgttcccttggtagaggtcgaatgcaaaacattggaaga 442  
QY 706 gccgcattcaattgcatgtgtagaagatcgaatcaataatgcaaatgagtggtgtc 765  
DB 443 gcaacttaagcgtcctaaggttggagatataaagcgaatcattgtgcatacagacgttgagtc 502  
QY 766 catactcccaagaacagagataa 789  
DB 503 cacatcataagaagaacagcttga 526

RESULT 4  
C36658  
ID C36658 standard; DNA: 835 BP.  
AC C36658;  
XX  
DT 17-OCT-2000 (first entry)  
XX

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 14609.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 11-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.

```

PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155466.
PR 24-SEP-1999; 99US-0155659.
PR 26-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156536.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160960.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 8.8%; Score 87.2; DB 21; Length 835;

Best Local Similarity 53.9%; Pred. No. 8.5e-13;

Matches 20; Conservative 0; Mismatches 168; Indels 9; Gaps 1;

```

OY 415 aaaaaaatgttcaaacgacattgtgaatgagatgagccaggttagatttcatttg 474
    || || || || || || || || || || || || || || || || || || ||
DB 145 aacaagacatcgaagaagagacagatgagtcgacgaagacgaattgaagcattc 204
    || || || || || || || || || || || || || || || || || || ||
OY 475 ttgtatcaagaacatcaatcaagaacacgaagtcgttgaaggtcccccagatgg 527
    || || || || || || || || || || || || || || || || || || ||
DB 205 ttgatcaagcctcaaggtctcagaggaagcgtctgtgaagatccacgaaggagaat 264
    || || || || || || || || || || || || || || || || || || ||
OY 528 --tataaataaactagacgaatctatacagatatgaagaacatttgaagaata 585
    || || || || || || || || || || || || || || || || || || ||
DB 265 attctgaactaccacttagagaagccgctcgagcagcttaaatcgatccgtgaagatatt 324
    || || || || || || || || || || || || || || || || || || ||
OY 586 ttgagtggtgaggttaactaagtgaattgccaataacgaagtgatctgacgtcacat 645
    || || || || || || || || || || || || || || || || || || ||
DB 325 gtcccgagcaagcggaatttcgaagaagtgcgagactcggttctcgtactgactcgct 384
    || || || || || || || || || || || || || || || || || || ||
OY 646 gaacgaggttggttgatttaggttttttagcaagaagacaaagcaccacattcgaaaga 705
    || || || || || || || || || || || || || || || || || || ||
DB 385 aaacgcgagcgtgactcttcttcttcttcttcttcttcttcttcttcttcttcttctt 444
    || || || || || || || || || || || || || || || || || || ||

```

```

OY 706 gccgcattcaatttcagatttgagagaagtcagtaacataattgaaccaatagtgtgtc 765
    || || || || || || || || || || || || || || || || || || ||
DB 445 gaacttacgacactcaaggtggagatataagcagatatgtctgacagacagtgagtc 504
    || || || || || || || || || || || || || || || || || || ||
OY 766 catatccctccaaagaagacagataa 789
    || || || || || || || || || || || || || || || || || || ||
DB 505 cacatcatcaagaagacagacttaa 528
    || || || || || || || || || || || || || || || || || || ||

```

RESULT 5  
T68888 standard; DNA; 1014 BP.

T68888; T68888;

09-DEC-1997 (first entry)

NIMA-interacting protein pin1 DNA.

pin1; protein interacting with NIMA; cell proliferation; mitosis;

peptidyl-prolyl cis/trans isomerase; adenocarcinoma; cancer;

leukemia; psoriasis; pemphigus vulgaris; rheumatoid arthritis;

acute respiratory distress syndrome; septic shock; inflammation;

therapy; ss.

Homo sapiens.

Key Location/Qualifiers

CDS 25..516

W0917986-A1.

22-MAY-1997.

28-OCT-1996; 96WO-US17334.

13-NOV-1995; 95US-0555912.

(SALK ) SALK INST BIOLOGICAL STUDIES.

Hunter T, Lu KP;

WPI; 1997-289057/26.

P-PSDB; W18312.

Protein, pin1, interacting with NIMA - used for treating cell

proliferative disorders

Claim 8; Page 49-50; 73pp; English.

This DNA sequence codes for human pin1 (W18312), an 18 kDa protein

that has peptidyl-prolyl cis/trans isomerase activity, associates

with NIMA protein kinase, inhibits the mitosis-promoting function

of NIMA when overexpressed, and induces mitotic arrest and nuclear

fragmentation when depleted. It was identified using a yeast two-

hybrid system with Aspergillus nidulans nima as bait and cDNA from

an HeLa library. A recombinant expression vector comprising the

DNA sequence and host cells containing the vector are claimed.

Methods are also claimed for identifying proteins that inhibit the

mitosis promoting function of NIMA protein kinase and for

controlling the growth of a cell by reducing pin1 activity or pin1

expression using an inhibitor, anti-pin1 antibody, antisense

nucleotide sequence or ribozyme, or by increasing pin1 activity

in the presence of an activator or increasing pin1 expression using

an enhancer. This allows treatment of cell proliferation disorders

such as adenocarcinomas, cancers, psoriasis, pemphigus vulgaris, acute

respiratory distress syndrome, rheumatoid arthritis, septic shock

and inflammation.

Sequence 1014 BP; 220 A; 310 C; 324 G; 160 T; 0 other;

Query Match 5.4%; Score 53.6; DB 18; Length 1014;  
Best Local Similarity 48.4%; Pred. No. 0.00026;  
Matches 180; Conservative 0; Mismatches 189; Indels 3; Gaps 1;

QY 418 aacatgttcaagccactgtgaatgagatggcgaagttagatgttcattgttg 477  
DB 148 agcagtggtgcaaaaacagggcaggggagccgtccagggctcgtccgacctgtg 207  
QY 478 atcaagaacaatcaatcaagaagaacccaagtcttggaagtcgccagatggtataga 537  
DB 208 gtgaagcagcagcagtcacgagccgtcgtccgtgagcagggagaagatcacccg 264  
QY 538 actagaagcgaatctatacagatattgaagaacatttggaagaataattgagtgtgag 597  
DB 265 accaagaagagagccctgagctgatacagcgtacatccagaagaatgctcgagag 324  
QY 598 gtaaacatgaattgtggcaataacccaagtattgtgacgtacatgaagaagtggt 657  
DB 325 gaggaacttgagctctgtgctccagctcagcagctgacatgaagcagggagga 384  
QY 658 gattagaggtttttagcaagaagaacccaatcaccacatcgaagaagccgcatc 717  
DB 385 gacctggtgctcttcagcagagaggtcagatgcagaagccatttgagaagcctgttg 444  
QY 718 ttgcattgtgagaagtcagtaacataattgaacccaatagtggttcatactccaa 777  
DB 445 ctgcgagcggggagatgagcgggcccggtgtcaagcagatccgcgacatcatcctc 504  
QY 778 agacagagataa 789  
DB 505 cgaactgagctga 516

RESULT 6  
F07816  
ID F07816 standard; cDNA; 662 BP.  
AC F07816;  
XX  
DT 13-MAR-2001 (first entry)  
XX  
DE Fusarium venenatum EST SEQ ID NO:339.  
XX  
KW Multiple gene expression; filamentous fungal cell; EST;  
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
KW culture condition; environmental stress; spore morphogenesis;  
KW metabolic pathway engineering; catabolic pathway engineering; ss.  
XX  
OS Fusarium venenatum.  
XX  
PN WO200056762-A2.  
XX  
PD 28-SEP-2000.  
XX  
PF 22-MAR-2000; 2000WO-US0781.  
XX  
PR 22-MAR-1999; 99US-0273623.  
XX  
PA (NOVO ) NOVO NORDISK INC.  
XX  
PI Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
XX  
DR WPI: 2000-594572/56.  
XX  
PT Monitoring differential expression of genes in filamentous fungal cells  
XX uses fluorescence-labeled nucleic acids isolated from the cells and a  
XX substrate of expressed sequence tags -  
XX  
PS Claim 86; Page 514; 3161pp; English.  
XX  
CC The present invention describes a method for monitoring differential

CC expression of genes in a first filamentous fungal (FF) cell relative to  
CC expression of the same genes in one or more second filamentous fungal  
CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
CC are used in the methods for monitoring differential expression of genes  
CC in a first filamentous fungal (FF) cell relative to expression of the  
CC same genes in one or more second filamentous fungal cells. Monitoring  
CC the global expression of genes from FF cells allows the production  
CC potential of the microorganisms to be improved. New genes may be  
CC discovered, possible functions of unknown open reading frames can be  
CC identified and gene copy number variation and stability can be  
CC monitored. The expression of genes can be used to study how FF cells  
CC adapt to changes in culture conditions, environmental stress, spore  
CC morphogenesis, recombination, metabolic or catabolic pathway engineering.  
CC Using ESTs provides several advantages over genomic or random cDNA  
CC clones including elimination of redundancy as one spot on an array  
CC equals one gene or open reading frame, and organization of the  
CC microarrays based on function of the gene products to facilitate  
CC analysis of the results. F07478 to F1247 represents ESTs from Fusarium  
CC venenatum; F11248 to F11853 represents ESTs from Aspergillus niger;  
CC F11854 to F14878 represents ESTs from Aspergillus oryzae; and F14879 to  
CC F15337 represents ESTs from Trichoderma reesei, which are all  
CC specifically claimed in the present invention.

XX  
SQ Sequence 662 BP; 169 A; 190 C; 159 G; 141 T; 3 other:

Query Match 5.4%; Score 53.2; DB 21; Length 662;  
Best Local Similarity 50.0%; Pred. No. 0.00029;  
Matches 248; Conservative 0; Mismatches 233; Indels 15; Gaps 4;

QY 271 acaggttaccacctaattgacgattagagatccagatccacataaagaatattc 330  
DB 119 accgtcttcctgctgtgctggaggtccgacactccacatccaaagccctccctattac 178  
QY 331 ttaaccaatctaccatgagtgctgttgaccacctatgagcactgcaagaagta 390  
DB 179 ttcaactcgcgaagaagctgtcgcgattggaacccctccgcagcagacccaagaag 238  
QY 391 ttgaatcatcattgcgaagttaaataaacaatggttacaaagccacttggatgagat 450  
DB 239 ctcaagcactacatgycgaacgaacataagcgtgttcgcgtcctgttgcaagttcccg 298  
QY 451 ggcacag-----ttagagttctcattgttgatcaagaacaatcaagaanaa 501  
DB 299 gtcccgagggcaaatctgcgtcccatctctcgtcaagcatcgcgacgacgcagc 358  
QY 502 cccaagcttggaagtcaccaagatggtataagttagactgagcagcatatcatacata 561  
DB 359 cccaacagctlgagagagccga---gattactgcgtccaagagagaggtcttcagatc 415  
QY 562 ttgaagaacattggaagaatattgagtgtgaggttaactaaagtaatt-tygcaaa 620  
DB 416 atcaagagcagcagaagaatcaagatctggaagtgtcagccttgcgaaagctgtctc 475  
QY 621 taccgaagtgtatgcagctcacatgcagaggtggtgattgaggtttttagcaa--a 678  
DB 476 caccgagctcgactgcctcctacatcgcagagcgttggtggtatttggaacttttgca 535  
QY 679 ggacaatgcaaccacatctcgagaagccgcatcattcaattgcatgttggaagaatcagt 738  
DB 536 ggagatatagcaaaaggttggaagatgcttcttcttcttcaagccacatgagc 595  
QY 739 aacataattgaacca 754  
DB 596 gagatgtcgaaacaa 611

RESULT 7  
F11336  
ID F11336 standard; cDNA; 406 BP.  
XX  
AC F11336;



XX 13-MAR-2001 (first entry)  
 DT  
 XX  
 DE Aspergillus niger EST SEQ ID NO:3859.  
 XX  
 KW Multiple gene expression; filamentous fungal cell; EST;  
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.  
 XX  
 OS Aspergillus niger.  
 XX  
 OS WO200056762-A2.  
 XX  
 XX  
 PD 28-SEP-2000.  
 XX  
 PF 22-MAR-2000; 2000WO-US07781.  
 XX  
 PR 22-MAR-1999; 99US-0273623.  
 XX  
 PA (NOVO ) NOVO NORDISK BIOTECH INC.  
 PA (NOVO ) NOVO NORDISK AS.  
 XX  
 PI Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
 XX  
 DR WPI: 2000-594572/56.  
 XX  
 PT Monitoring differential expression of genes in filamentous fungal cells  
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags -  
 XX  
 PS Claim 87; Page 1725; 3161pp; English.  
 XX  
 CC The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring  
 CC the global expression of genes from FF cells allows the production  
 CC potential of the microorganisms to be improved. New genes may be  
 CC discovered, possible functions of unknown open reading frames can be  
 CC identified and gene copy number variation and stability can be  
 CC monitored. The expression of genes can be used to study how FF cells  
 CC adapt to changes in culture conditions, environmental stress, spore  
 CC morphogenesis, recombination, metabolic or catabolic pathway engineering.  
 CC Using ESTs provides several advantages over genomic or random cDNA  
 CC clones including elimination of redundancy as one spot on an array  
 CC equals one gene or open reading frame, and organisation of the  
 CC microarrays based on function of the gene products to facilitate  
 CC analysis of the results. F07478 to F11247 represents ESTs from Fusarium  
 CC venenatum; F11248 to F11853 represents ESTs from Aspergillus niger;  
 CC F11854 to F14878 represents ESTs from Aspergillus oryzae; and F14879 to  
 CC F15337 represents ESTs from Trichoderma reesei, which are all  
 CC specifically claimed in the present invention.  
 XX  
 SO Sequence 406 BP; 97 A; 86 C; 124 G; 93 T; 6 other;

Query Match 5.3%; Score 52.4; DB 21; Length 406;  
 Best Local Similarity 49.6%; Pred. No. 0.00039;  
 Matches 134; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

OY 522 agatgtataagtagactagacgaactatatacagatatgaagaacattggaag 581  
 II II II  
 DB 69 agctgaatcacccgttcgaaagagagcgcgcgacgtcccgcgacgaagcgcg 128  
 II II II  
 OY 562 aatattggtgtgtggttaactaagtgaattgccaataacgcgaagaagtgcagccc 641  
 II II II  
 DB 129 catcaatgcgggaagaccagctctgtgtagcatctgtatcgcgagtcgactgagcag 188  
 II II II I

OY 642 acatgcagagtggtgattaggttttttagcaagaagcaaatgcacaccacattcga 701  
 I  
 DB 189 tgcacaggaacagctgtagattggttcttcttgacggaattgtagtcagaagaagttaga 248  
 II II II  
 OY 702 agaagccgcattcaatttcattgttggagagtcagtaacataattgaaaccaattg 761  
 II II II  
 DB 249 ggacgcgcctctgtcacctcacccgcgtcagtcagtcagtggtgtagagacgcctctg 308  
 II II II  
 OY 762 tgtccatccctcccaagaacaggaataat 791  
 II II II  
 DB 309 tgtccattgattgaacggttcaagtaaat 338  
 II II II I

RESULT 8  
 V74479  
 ID V74479 standard; DNA; 7963 BP.  
 XX  
 AC V74479;  
 XX  
 XX  
 DT 16-MAR-1999 (first entry)  
 XX  
 DE Staphylococcus aureus contig SEQ ID #168.  
 XX  
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.  
 XX  
 OS Staphylococcus aureus.  
 XX  
 OS  
 XX  
 FH Key location/Qualifiers  
 FT misc\_feature 1501..1560  
 FT /tag= a  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT  
 FT misc\_feature 3301..3360  
 FT /tag= b  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT  
 FT misc\_feature 5101..5160  
 FT /tag= c  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT  
 FT misc\_feature 6901..6960  
 FT /tag= d  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT  
 PN EP786519-A2.  
 XX  
 PD 30-JUL-1997.  
 XX  
 XX  
 PE 07-JAN-1997; 97EP-0100117.  
 XX  
 PR 05-JAN-1996; 96US-0009861.  
 XX  
 PA (HUMA-) HUMA GENOME SCT INC.  
 PA Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
 PI Rosen CA;  
 XX  
 DR WPI: 1997-374922/35.  
 XX  
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus



DE Arabidopsis thaliana DNA fragment SEQ ID NO: 9547.  
XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
XX metabolic pathway; promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
XX EPI033405-A2.  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 23-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136382.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139482.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146389.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.

PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159395.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 3.9%; Score 38.6; DB 21; Length 666;  
Best Local Similarity 52.1%; Pred. No. 1.4;  
Matches 86; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
QY 630 tgaatgcagctacatcacagaggtgtgatttaagggttttttagcaagaagacaatgca 689  
DB 343 ttaggtgtccatccagggaaaaagagagatcttgaatgtgtcccttgaggaaagatgac 402  
QY 690 accaccatttgaagaagccgcatcatttgcattgttggagagtcagttacataatgca 749  
DB 403 aggtccattccaaagagtgcccttaatacacctgttggaggtcacccgtgcaccccttcaa 462  
QY 750 aaccaatagtggtgtcatatcctccaagaagaacaggaataatcaa 794  
DB 463 atcaagcagcagatcacacatttattatcagagggagaagaaga 507

RESULT 11  
X20248 standard; DNA; 910715 BP.  
AC X20248;  
XX  
DT 04-MAY-1999 (first entry)

XX  
DE Borrelia burgdorferi polynucleotide sequence #1.  
XX  
KW Borrelia burgdorferi: spirochete; bacterium; pathogen; Lyme disease;  
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;  
KW infection; diagnosis; characterisation; detection; ds.  
XX  
OS Borrelia burgdorferi.  
XX  
PN W09858943-A1.  
XX  
PD 30-DEC-1998.  
XX  
PF 18-JUN-1998; 98WO-US12764.  
XX  
PR 03-SEP-1997; 97US-0057483.  
PR 20-JUN-1997; 97US-0050359.  
PR 22-JUL-1997; 97US-0053344.  
PR 22-JUL-1997; 97US-0053377.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PA (MED1-) MEDIMUNE INC.  
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;  
PI White OR;  
XX  
DR WPI; 1999-081217/07.  
XX  
PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
PT products for the detection, diagnosis, characterisation, prevention  
PT and therapy of infections, particularly Lyme disease  
XX  
PS Claim 1; Page 157-671; 1128pp; English.  
XX  
CC X20248 to X20402 represent polynucleotide sequences isolated from  
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for  
CC the detection, diagnosis, characterisation, prevention and therapy of  
CC Bb infections, e.g. Lyme disease. They can also be used for the  
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs  
CC to a family of motile, spiral-shaped bacteria called Spirochetes.  
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and  
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as  
CC Lyme disease.  
XX  
SQ Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T; 54 other;

Query Match 3.9%; Score 38.4; DB 20; Length 910715;  
Best Local Similarity 52.5%; Pred. No. 15;  
Matches 84; Conservative 0; Mismatches 76; Indels 0; Gaps 0;  
QY 701 aagaagccgcatcatttgcattgttgagagatcagtaacataatgaaacaaatgtg 760  
DB 383284 aaataagactcaactaaatgtctctgtatctacagagcatcatttaattttgaaatcaaaagac 383343  
QY 761 gtgtccatctccccaagaacagataaatacaagatattggaattgttgataaatagaa 820  
DB 383344 aattaatctactcttaagaatataatataataaactcatttaaaaaaacaacaaca 383403  
QY 821 aataaataagagacaagttgtatagattgtgtacccaanaa 860  
DB 383404 tcttaataaaaactattttaaaaatttcctaactaca 383443

RESULT 12  
C62126 standard; DNA; 4468 BP.  
ID C62126  
XX  
AC C62126;  
XX  
DT 06-MAR-2001 (first entry)  
XX  
DE DNA encoding a calcium-dependent serine-protease designated Pf-SUB2.

XX Calcium-dependent serine-protease; Pf-SUB2; merozoite differentiation;  
 KW major surface protein 1; MSP1-42; erythrocyte entry; malaria; ss.  
 XX Plasmodium falciparum.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 61..4074  
 FT /\*tag= a  
 FT /product= "serine protease"  
 XX  
 PN FR2791685-A1.  
 PD 06-OCT-2000.  
 XX  
 PF 31-MAR-1999; 99FR-0004039.  
 XX  
 PR 31-MAR-1999; 99FR-0004039.  
 XX  
 PA (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Barale JC, Langsley G, Braun BC, Pereira Da Silva L, Blisnick T;  
 XX WPI: 2000-658021/64.  
 DR P-PSDB; B30504.  
 XX  
 PT Polypeptide with calcium-dependent serine-protease activity, for the  
 PT prevention, treatment, and detection of malarial infections due to  
 PT Plasmodium falciparum  
 XX  
 PS Claim 3; Page 33-34; 47pp; French.  
 XX  
 CC The present sequence encodes a polypeptide (Pf-SUB2) which has a  
 CC calcium-dependent serine-protease activity. The Pf-SUB2 gene is  
 CC expressed during the differentiation phase of merozoites. The protein  
 CC is implicated in maturation of the major surface protein 1 of  
 CC merozoites (MSP1-42). The enzyme is also crucial for entry of the  
 CC parasite into erythrocytes. The polypeptides and polynucleotides are  
 CC used to identify inhibitors of Pf-SUB2. These inhibitors e.g. antibodies,  
 CC are used for the detection, prevention, and treatment of malaria due to  
 CC Plasmodium falciparum infection.  
 CC  
 XX  
 SQ Sequence 4468 BP; 1902 A; 456 C; 641 G; 1469 T; 0 other;  
 Query Match 3.9%; Score 38.2; DB 21; Length 4468;  
 Best Local Similarity 52.1%; Pred. No. 3.2;  
 Matches 85; Conservative 0; Mismatches 78; Indels 0; Gaps 0;  
 Oy 74 attagcactggaatagatatagattgagattgcttgcgaatagatataggtagtg 133  
 |||| | ||| |||| | || | || |||| | |||  
 Db 4248 attatataatataatataatataatataatataatataatataatattagatat 4307  
 Oy 134 tacattcaaaaactctctcttttctatattcttcaatcaacaagaatttcgtgt 193  
 |||| | ||| || |||| | || | |||| | |||  
 Db 4308 ttatttttaaatagtaactctttttatgtgaaacaacattccctcttttcgtgt 4367  
 Oy 194 tgcctttgtgtattatttgcctcagttcagttgattctt 236  
 |||| | |||| | |||| | |||| | |||| | ||||  
 Db 4368 ttcatattattattattattattattattattttttttt 4410  
 RESULT 13  
 C62127  
 ID C62127 standard; DNA; 4611 BP.  
 XX  
 AC C62127;  
 XX  
 DT 06-MAR-2001 (first entry)  
 XX  
 DE DNA encoding a calcium-dependent serine-protease designated Pf-SUB2.  
 XX

KW Calcium-dependent serine-protease; Pf-SUB2; merozoite differentiation;  
 KW major surface protein 1; MSP1-42; erythrocyte entry; malaria; ss.  
 XX Plasmodium falciparum.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 124..4217  
 FT /\*tag= a  
 FT /product= "serine protease"  
 FT /note= "contains an intron"  
 FT exon 124..2877  
 FT /\*tag= b  
 FT Intron 2878..2930  
 FT /\*tag= c  
 FT exon 2931..4217  
 FT /\*tag= d  
 XX  
 PN FR2791685-A1.  
 PD 06-OCT-2000.  
 XX  
 PF 31-MAR-1999; 99FR-0004039.  
 XX  
 PR 31-MAR-1999; 99FR-0004039.  
 XX  
 PA (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Barale JC, Langsley G, Braun BC, Pereira Da Silva L, Blisnick T;  
 XX WPI: 2000-658021/64.  
 DR P-PSDB; B30505.  
 XX  
 PT Polypeptide with calcium-dependent serine-protease activity, for the  
 PT prevention, treatment, and detection of malarial infections due to  
 PT Plasmodium falciparum  
 XX  
 PS Claim 3; Page 34-45; 47pp; French.  
 XX  
 CC The present sequence encodes a polypeptide (Pf-SUB2) which has a  
 CC calcium-dependent serine-protease activity. The Pf-SUB2 gene is  
 CC expressed during the differentiation phase of merozoites. The protein  
 CC is implicated in maturation of the major surface protein 1 of  
 CC merozoites (MSP1-42). The enzyme is also crucial for entry of the  
 CC parasite into erythrocytes. The polypeptides and polynucleotides are  
 CC used to identify inhibitors of Pf-SUB2. These inhibitors e.g. antibodies,  
 CC are used for the detection, prevention, and treatment of malaria due to  
 CC Plasmodium falciparum infection.  
 CC  
 XX  
 SQ Sequence 4611 BP; 1951 A; 459 C; 657 G; 1544 T; 0 other;  
 Query Match 3.9%; Score 38.2; DB 21; Length 4611;  
 Best Local Similarity 52.1%; Pred. No. 3.2;  
 Matches 85; Conservative 0; Mismatches 78; Indels 0; Gaps 0;  
 Oy 74 attagcactggaatagatatagattgagattgcttgcgaatagatataggtagtg 133  
 |||| | ||| |||| | || | || |||| | |||  
 Db 4391 attatataatataatataatataatataatataatataatattagatat 4450  
 Oy 134 tacattcaaaaactctcttttctatattcttcaatcaacaagaatttcgtgt 193  
 |||| | ||| || |||| | || | |||| | |||  
 Db 4451 ttatttttaaatagtaactctttttatgtgaaacaacattccctcttttcgtgt 4510  
 Oy 194 tgcctttgtgtattatttgcctcagttcagttgattctt 236  
 |||| | |||| | |||| | |||| | |||| | ||||  
 Db 4511 ttcatattattattattattattattattttttttt 4553  
 RESULT 14  
 Z01425  
 ID Z01425 standard; DNA; 1038602 BP.  
 XX



Oy	209	tattgcatcagtttagcttgatctcttcttgacgtatgatcalcatcgactgcacat	268
Db	675	TTTTTTTCTTTTTTTTTTTTTTTTTTGCGAAAAAAATGCAAGTGTATTATGA	616
Oy	269	caacaggcttacccaacctaatigacgagttagaghatccagaaccocataaacaaagaglat	328
Db	615	AGCCAGTGGTAAAAGTAAATAAAGGCCCTTTTGCCTTTTCATTTCAATTTCCAAGAAGAGAT	556
Oy	329	tcttaaaccaa tctaccaa tga gtcg cgt ctggg acc ca act ta gca	376
Db	555	TATCTTCAAAAGACCCCAAGTGTAAGTAGTGCCAGAACATAAAAAGCA	508

Search completed: June 8, 2001, 06:21:22  
Job time: 6146 sec

**This Page Blank (uspto)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2001, 03:58:30 ; Search time 1747.82 Seconds  
(without alignments)  
4943.301 Million cell updates/sec

Title: US-09-507-242-1  
Perfect score: 989  
Sequence: 1 gatcaacacatagatgtgt.....ggtttgtcatatttgatcc 989

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_esthum29:\*  
64: em\_esthum30:\*  
65: em\_esthum31:\*  
66: em\_esthum32:\*  
67: em\_esthum33:\*  
68: em\_esthum34:\*  
69: em\_esthum35:\*  
70: em\_esthum36:\*  
71: em\_esthum37:\*  
72: em\_esthum38:\*  
73: em\_esthum39:\*  
74: em\_esthum40:\*  
75: em\_esthum41:\*  
76: em\_esthum42:\*  
77: em\_esthum43:\*  
78: em\_esthum44:\*  
79: em\_esthum45:\*  
80: em\_esthum46:\*  
81: em\_esthum47:\*  
82: em\_esthum48:\*  
83: em\_esthum49:\*  
84: em\_esthum50:\*  
85: em\_esthum51:\*  
86: em\_esthum52:\*  
87: em\_esthum53:\*  
88: em\_esthum54:\*  
89: em\_esthum55:\*  
90: em\_esthum56:\*  
91: em\_esthum57:\*  
92: em\_esthum58:\*  
93: em\_esthum59:\*  
94: em\_esthum60:\*  
95: em\_esthum61:\*  
96: em\_esthum62:\*  
97: em\_esthum63:\*  
98: em\_esthum64:\*  
99: em\_esthum65:\*  
100: em\_esthum66:\*  
101: em\_esthum67:\*  
102: em\_esthum68:\*  
103: em\_esthum69:\*  
104: em\_esthum70:\*  
105: em\_esthum71:\*  
106: em\_esthum72:\*  
107: em\_esthum73:\*  
108: em\_esthum74:\*  
109: em\_esthum75:\*  
110: em\_esthum76:\*  
111: em\_esthum77:\*  
112: em\_esthum78:\*  
113: em\_esthum79:\*  
114: em\_esthum80:\*  
115: em\_esthum81:\*  
116: em\_esthum82:\*

117: gb\_est48:\*  
 118: gb\_est49:\*  
 119: gb\_est50:\*  
 120: gb\_est51:\*  
 121: gb\_est52:\*  
 122: gb\_est53:\*  
 123: gb\_est54:\*  
 124: gb\_est55:\*  
 125: gb\_est56:\*  
 126: gb\_est57:\*  
 127: gb\_est58:\*  
 128: gb\_est59:\*  
 129: gb\_est60:\*  
 130: gb\_est61:\*  
 131: gb\_est62:\*  
 132: gb\_est63:\*  
 133: gb\_est64:\*  
 134: gb\_est65:\*  
 135: gb\_est66:\*  
 136: gb\_est67:\*  
 137: gb\_est68:\*  
 138: gb\_est69:\*  
 139: gb\_est70:\*  
 140: gb\_est71:\*  
 141: gb\_est72:\*  
 142: gb\_est73:\*  
 143: gb\_est74:\*  
 144: gb\_est75:\*  
 145: gb\_est76:\*  
 146: gb\_est77:\*  
 147: gb\_est78:\*  
 148: gb\_est79:\*  
 149: gb\_est80:\*  
 150: gb\_est81:\*  
 151: gb\_est82:\*  
 152: gb\_est83:\*  
 153: gb\_est84:\*  
 154: gb\_est85:\*  
 155: gb\_est86:\*  
 156: gb\_est87:\*  
 157: gb\_est88:\*  
 158: gb\_est89:\*  
 159: gb\_est90:\*  
 160: gb\_est91:\*  
 161: gb\_est92:\*  
 162: gb\_est93:\*  
 163: gb\_est94:\*  
 164: gb\_est95:\*  
 165: gb\_est96:\*  
 166: gb\_est97:\*  
 167: gb\_est98:\*  
 168: gb\_est99:\*  
 169: gb\_est100:\*  
 170: gb\_est101:\*  
 171: gb\_est102:\*  
 172: gb\_est103:\*  
 173: gb\_est104:\*  
 174: gb\_est105:\*  
 175: gb\_est106:\*  
 176: gb\_est107:\*  
 177: gb\_est108:\*  
 178: gb\_est109:\*  
 179: gb\_est110:\*  
 180: gb\_est111:\*  
 181: gb\_est112:\*  
 182: gb\_est113:\*  
 183: gb\_est114:\*  
 184: gb\_est115:\*  
 185: gb\_est116:\*  
 186: gb\_est117:\*  
 187: gb\_est118:\*  
 188: gb\_est119:\*  
 189: gb\_est120:\*

190: em\_gss\_pln1:\*  
 191: em\_gss\_pln2:\*  
 192: em\_gss\_pro:\*  
 193: em\_gss\_rod1:\*  
 194: em\_gss\_rod2:\*  
 195: em\_gss\_rod3:\*  
 196: em\_gss\_rod4:\*  
 197: em\_gss\_rod5:\*  
 198: em\_gss\_vrt1:\*  
 199: em\_gss\_vrt2:\*  
 200: em\_gss\_vrt3:\*  
 201: gb\_gss1:\*  
 202: gb\_gss2:\*  
 203: gb\_gss3:\*  
 204: gb\_gss4:\*  
 205: gb\_gss5:\*  
 206: gb\_gss6:\*  
 207: gb\_gss7:\*  
 208: gb\_gss8:\*  
 209: gb\_gss9:\*  
 210: gb\_gss10:\*  
 211: gb\_gss11:\*  
 212: gb\_gss12:\*  
 213: gb\_gss13:\*  
 214: gb\_gss14:\*  
 215: gb\_gss15:\*  
 216: gb\_gss16:\*  
 217: gb\_gss17:\*  
 218: gb\_gss18:\*  
 219: gb\_gss19:\*  
 220: gb\_gss20:\*  
 221: gb\_gss21:\*  
 222: gb\_gss22:\*  
 223: gb\_gss23:\*  
 224: gb\_gss24:\*  
 225: gb\_gss25:\*  
 226: gb\_gss26:\*  
 227: gb\_gss27:\*  
 228: gb\_gss28:\*  
 229: gb\_gss29:\*  
 230: gb\_gss30:\*  
 231: gb\_gss31:\*  
 232: gb\_gss32:\*  
 233: gb\_gss33:\*  
 234: gb\_gss34:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	101.2	10.2	569	121	AM669911 NXNV_123-
2	100.2	10.1	480	138	BE662291 ST84/ST84
3	99.4	10.1	601	110	AM064606 ST33E10 P
4	96.4	9.7	412	32	AM010930 ST13C12 P
5	96	9.7	615	114	AM334625 S3BD12 AG
6	94.8	9.6	462	137	BE607245 NXCT_034-
7	94.2	9.5	471	106	AU051808 AU051808
8	94.2	9.5	517	175	C92940 C92940 Dict
9	94.2	9.5	659	175	C92104 C92104 Dict
10	92	9.3	729	106	AU000273 AU000273
11	90.6	9.2	708	32	AM010527 ST07G11 P
12	90.4	9.1	738	29	AV400559 AV400559
13	90.4	9.1	748	106	AU003598 AU003598
14	89.8	9.1	660	106	AU000274 AU000274
15	89.8	9.1	722	106	AU003857 AU003857
16	89.8	9.1	731	106	AU003742 AU003742
17	89.8	9.1	740	106	AU003792 AU003792
18	89.6	9.1	630	175	C90182 C90182 Dict

```

19 87.2 8.8 703 162 BE038350
20 81.8 8.3 577 24 A1772554
21 81.8 8.3 577 173 BG133271
22 81.8 8.3 648 122 AM945046
23 80.6 8.1 699 102 A1822163
24 80.6 8.1 780 161 BE034276
25 76.4 7.7 409 149 BF519916
26 76.4 7.7 492 152 BF519382
27 76.4 7.7 587 142 BF003579
28 76 7.7 618 114 AM397670
29 76 7.7 671 114 AM348975
30 73.2 7.4 665 110 AM054423
31 73 7.4 437 21 A1496600
32 72 7.3 356 157 T76444
33 72 7.3 810 146 BF280096
34 71.8 7.3 467 165 BE346524
35 71.8 7.3 511 120 AM831738
36 71.4 7.2 497 165 BE316431
37 70.8 7.2 513 105 AL387506
38 70.4 7.2 205 111 AM145706
39 70.4 7.1 960 162 BE054767
40 70.2 7.1 452 163 BE190190
41 70.2 7.1 503 146 BF324477
42 70 7.1 261 141 BE921583
43 69.8 7.1 702 24 A1730273
44 68 6.9 400 119 AM761425
45 67.8 6.9 642 162 BE037023

```

## ALIGNMENTS

```

RESULT 1
LOCUS AM869911 569 bp mRNA EST 22-MAY-2000
DEFINITION NXNV_123_A01.F NSF Xylem Normal wood Vertical Pinus taeda cDNA
ACCESSION AM869911
VERSION AM869911.1 GI:8004060
KEYWORDS EST.
SOURCE loblobolly pine.
ORGANISM Pinus taeda
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE 1 (bases 1 to 569)
AUTHORS Sederoff, R.
TITLE Molecular Basis of Wood Formation in the Pine Megagenome
JOURNAL Unpublished (2000)
COMMENT Contact: Johnson, Arthur
North Carolina State University
Tel: 919 515 7800
Fax: 919 515 7801
Email: ajohnson@unity.ncsu.edu
Seq primer: T3.
FEATURES
source
1..569
Location/Qualifiers
/organism="Pinus taeda"
/db_xref="taxon:3352"
/clone="NXNV_123_A01"
/clone_lib="NSF Xylem Normal wood Vertical"
/note="Vector: Bluescript SK; Site_1: Eco RI; the
sequences contain a 'CDNA adapter' between the EcoRI site
and the start of the EST. The adapter sequence is
'AATTGGCAGCAGC'."
BASE COUNT 172 a 105 c 131 g 140 t 21 others
ORIGIN

```

Query Match 10.2%; Score 101.2; DB 121; Length 569;  
 Best Local Similarity 58.1%; Pred No. 2.4e-14;  
 Matches 200; Conservative 0; Mismatches 135; Indels 9; Gaps 1;

Qy 455 aggttagagttctcatctgttgatcaagaacatcaatcaagaacccaagcttctga 514

```

Db 127 AGGTGAGAGCTTCTCACTTGCTTATCAACAGAGGATCAAGAGGAGCTTCTCTGTC 186
515 agtcccaagatgtataag-----tagaactagaagaacatctacagatatga 565
187 AAGACCCAGATGCGCCAGAGATTAAAGCTAACCCAGAGATGCTGCGTGGCCCACTCA 246
566 agaaacatttggaagaacatcttgatggtgaggttaactaagtgaattggcaaatccg 625
247 GTCATTTGAGGAGAGAAATTTGTTAGTGAAGACCAAGTTTAGACCTTGGCGCTCGAT 306
626 aaagtatgtcagctcacatgacagaggtggtatgaagtttttaagcaagaacaa 685
307 ATTCGACGTGTAATTCGCTTAAGAAAGCGGAGATTGGTCCATTTGGTCGAGTCGAA 366
686 tgcacaccatttgcagaagacgcgcatctcaatttgatgttgagagtcagtaataa 745
367 TGCAGAAAGCCTTTCGAAAGTGTCTACATACCTTTGAAATGCGCAGATTGATGATTTG 426
746 ttgaacacatagtggtgtccatcatccccaagaacagatata 789
427 TGCACACCGACAGTGTGTGATATCATCTTAANNACAGATGA 470

```

```

RESULT 2
LOCUS BE662291 480 bp mRNA EST 08-SEP-2000
DEFINITION ST84/ST84C02 Pine Triplex shoot tip library Pinus taeda cDNA clone
ACCESSION BE662291
VERSION BE662291.1 GI:9993028
KEYWORDS EST.
SOURCE loblobolly pine.
ORGANISM Pinus taeda
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE 1 (bases 1 to 480)
AUTHORS Whetten, R.W., Kinlaw, C.S., Retzel, E. and Sederoff, R.R.
TITLE The Pine Gene Discovery Project
JOURNAL Unpublished (1999)
COMMENT Contact: Ross Whetten
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhetten@unity.ncsu.edu
Seq primer: 5' lambda triplex2 Sequencing Primer.
FEATURES
source
1..480
Location/Qualifiers
/organism="Pinus taeda"
/db_xref="taxon:3352"
/clone="ST84/ST84C02"
/clone_lib="Pine Triplex shoot tip library"
/lab_host="E. coli BM25.8"
/note="Organ: shoot tips; Vector: Lambda Triplex; Site_1:
SfiI (A); Site_2: SfiI (B); Shoot tips (approx. 2 cm from
apex) were collected during the spring, frozen and used
for mRNA isolation. The SMART-PCR method (Clontech) was
used to prepare a library from 1 ug total RNA, using the
Lambda Triplex vector. Plasmid subclones in pTriplex were
recovered by cre-lox excision in E. coli strain BM25.8 and
sequenced from the 5' end."
BASE COUNT 138 a 85 c 119 g 122 t 16 others
ORIGIN

```

Query Match 10.1%; Score 100.2; DB 138; Length 480;  
 Best Local Similarity 57.8%; Pred No. 4.1e-14;  
 Matches 199; Conservative 0; Mismatches 136; Indels 9; Gaps 1;

Qy 455 aggttagagttctcatctgttgatcaagaacatcaatcaagaacccaagcttctga 514

```

Db 38 AGGTGAGAGCTTCTACACTTGGCTTATCAACACGAGGATCAAGAGGCGCTTCCTCGTGGC 97
Qy 515 agtccagatggtataga-----tagaactagagaactctatacagatatga 565
Db 98 AAGACCCAGATGCGCGAAGGATTAAGCTACACACAGATGCTGCCGTTCCCACTCA 157
Qy 566 agaaacattggaagaatattgagtggtgaggttaactaagtgaaattggaataccg 625
Db 158 GTGATTGAGGAGGAATTTAGTGAAGGCCAAGTTGAGGACCTTCGGCTCGAT 217
Qy 626 aaagtattgacgtccacatgacagaggtggtgatttagggttttttagcaagagacaa 685
Db 218 ATTCGACTGTAAATCCGCTTAAGAAAGGNGNATTTGGGTCATTGTCGACGTCAAA 277
Qy 686 tgcacacacattcgaaagccgcatcattgcatgtgttgaggaagtgacataa 745
Db 278 TGCAAAAGCCTTTGGAAGATCTCTACATACCTTTGAAAGTTGGCGAGATTGATGATTTG 337
Qy 746 ttgaaccaatagtggtgtccatatacctccaagaagacagataa 789
Db 338 TGGACACTGACAGTGTGTGCATATCATCTTAAGACAGAGATGA 381

```

```

RESULT 3
AM064606 601 bp mRNA EST 12-OCT-1999
LOCUS ST33E10, Pine Triplex shoot tip library Pinus taeda cDNA clone
DEFINITION ST33E10, mRNA sequence.
ACCESSION AM064606
VERSION AM064606.1 GI:6019678
KEYWORDS EST.
SOURCE loblolly pine.
ORGANISM Pinus taeda
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.

```

```

REFERENCE 1 (bases 1 to 601)
AUTHORS Whetten,R.W., Kinlaw,C.S., Retzel,E. and Sederoff,R.R.
TITLE The Pine Gene Discovery Project
JOURNAL Unpublished (1999)
COMMENT Contact: Ross Whetten
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
,NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhet@unity.ncsu.edu
Seq primer: 5' lambda Triplex2 Sequencing Primer.

```

## FEATURES

## source

```

1.601
/organism="Pinus taeda"
/db_xref="taxon:3352"
/clone="SP33E10"
/clone_1b="Pine Triplex shoot tip library"
/lab_host="E. coli BM25.8"
/note="Organ: shoot tips; Vector: Lambda Triplex; Site_1:
SfiI (A); Site_2: SfiI (B); Shoot tips (approx. 2 cm from
apex) were collected during the spring, frozen and used
for mRNA isolation. The SMART-PCR method (Clontech) was
used to prepare a library from 1 ug total RNA, using the
Lambda Triplex vector. Plasmid subclones in pTriplex were
recovered by cre-lox excision in E. coli strain BM25.8 and
sequenced from the 5' end."
BASE COUNT 189 a 113 c 147 g 150 t 2 others
ORIGIN

```

```

Query Match 10.1%; Score 99.4; DB 110; Length 601;
Best Local Similarity 56.9%; Pred. No. 6.6e-14;
Matches 206; Conservative 0; Mismatches 147; Indels 9; Gaps 1;

```

```

Qy 455 aggttagagttctccattgtgatcaagaacatcaatcaagaagcagtccttga 514

```

```

Db 160 AGGTGAGAGCTTCTACACTTGGCTTATCAACACGAGGATCAAGAGGCGCTTCCTCGTGGC 219
Qy 515 agtccagatggtataga-----tagaactagagaactctatacagatatga 565
Db 220 AAGACCCAGATGCGCGAAGGATTAAGCTACACACAGATGCTGCCGTTCCCACTCA 279
Qy 566 agaaacattggaagaatattgagtggtgaggttaactaagtgaaattggaataccg 625
Db 280 GTGATTGAGGAGGAATTTAGTGAAGGCCAAGTTGAGGACCTTCGGCTCGAT 339
Qy 626 aaagtattgacgtccacatgacagaggtggtgatttagggttttttagcaagagacaa 685
Db 340 ATTCGACTGTAAATCCGCTTAAGAAAGGCGGAGATTTGGGTCATTGTCGACGTCAAA 399
Qy 686 tgcacacacattcgaaagccgcatcattgcatgtgttgaggaagtgacataa 745
Db 400 TGCAAAAGCCTTTGGAAGATCTCTACATACCTTTGAAAGTTGGCGAGATTGATGATTTG 459
Qy 746 ttgaaccaatagtggtgtccatatacctccaagaagacagataaataagattgagt 805
Db 460 TGGACACTGACAGTGTGTGCATATCATCTTAAGACAGAGATGA 519
Qy 806 tt 807
Db 520 TT 521

```

```

RESULT 4
AM010930 412 bp mRNA EST 10-SEP-1999
LOCUS ST13C12, Pine triplex shoot tip library Pinus taeda cDNA clone
DEFINITION ST13C12, mRNA sequence.
ACCESSION AM010930
VERSION AM010930.1 GI:5859624
KEYWORDS EST.
SOURCE loblolly pine.
ORGANISM Pinus taeda
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.

```

```

REFERENCE 1 (bases 1 to 412)
AUTHORS Whetten,R.W., Kinlaw,C.S., Retzel,E. and Sederoff,R.R.
TITLE The Pine Gene Discovery Project
JOURNAL Unpublished (1999)
COMMENT Contact: Ross Whetten
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
,NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhet@unity.ncsu.edu
Seq primer: 5' lambda Triplex2 Sequencing Primer.

```

## FEATURES

## source

```

1.412
/organism="Pinus taeda"
/db_xref="taxon:3352"
/clone="SP13C12"
/clone_1b="Pine Triplex shoot tip library"
/lab_host="E. coli BM25.8"
/note="Organ: shoot tips; Vector: Lambda Triplex; Site_1:
SfiI (A); Site_2: SfiI (B); Shoot tips (approx. 2 cm from
apex) were collected during the spring, frozen and used
for mRNA isolation. The SMART-PCR method (Clontech) was
used to prepare a library from 1 ug total RNA, using the
Lambda Triplex vector. Plasmid subclones in pTriplex were
recovered by cre-lox excision in E. coli strain BM25.8 and
sequenced from the 5' end."
BASE COUNT 115 a 82 c 107 g 104 t 4 others
ORIGIN

```

```

Query Match 9.7%; Score 96.4; DB 32; Length 412;

```



BASE COUNT 141 a 89 c 115 g 107 t 10 others  
 ORIGIN

Query Match 9.68; Score 94.8; DB 137; Length 462;  
 Best Local Similarity 56.18; Pred. No. 8.3e-13;  
 Matches 193; Conservative 0; Mismatches 142; Indels 9; Gaps 1;

455 aggttagatcttcattcattttagcaagaacaatcaatgaagaagcccaagcttga 514  
 119 AGGCAAGAGCTTCTCATTGCTTATCAAAACACGAGGATCAAGAGGCTTCTGTCG 178  
 515 agtcccaagatgtgtataa-----tagaactagaacgaatctatacagattga 565  
 179 AAGACCCAGATGGCCGAAGATTAAGCTACACACAGATGCTGCCGTTGCCACTCA 238  
 566 agaacaatttggaaagatatattgagtgaggttaactaagtgaattggcaaatccg 625  
 239 GTGCATTGAGGAGGAATGTTAGTGAAGGCCCAAGTTGAGACCTTGCCTCGAT 298  
 626 aaagttagtgcagctacatgacagaggtgtgtattgaaggttttttagcaagaagaca 685  
 299 ATTCCGACTGTGTAATCCGCTTNCAGAACNAGATTGGGCTTCATTGGTGANNTC 358  
 686 tgcacaccacattcgaaagccgcatcattgtgcatgttggagagtcagtaacata 745  
 359 TGCNAAGCCCTTTGAGAGATGCTACATACCTTTTGAAGAAGTGGGAGATTAGTAT 418  
 746 ttgaaccataatgtgtgtcatatctctccaagaagaagataa 789  
 419 NNGACACTGACAGTGTGTGATTCATCTTAAGACAGAGATGA 462

RESULT 7  
 LOCUS AU051808 471 bp mRNA EST 26-MAR-1999  
 DEFINITION AU051808 Dictyostelium discoidium SS (H.Urushihara) Dictyostelium  
 accession AU051808  
 version AU051808.1 GI:4526753  
 keywords EST.  
 source Dictyostelium discoidium.  
 organism Dictyostelium discoidium.  
 reference Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
 authors 1 (bases 1 to 471)  
 title Yoshino, R., Morio, T. and Tanaka, Y.  
 journal Developmental cDNA in Dictyostelium discoidium  
 comment Unpublished (1997)  
 contact: Hideko Urushihara  
 institute of Biological Sciences  
 University of Tsukuba  
 3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan  
 email: d402huesakura.cc.tsukuba.ac.jp  
 product = 'Dictyostelium discoidium cDNA project in Japan'  
 location/Qualifiers

FEATURES  
 source  
 1. 471  
 /organism="Dictyostelium discoidium"  
 /db\_xref="taxon:44689"  
 /clone\_lib="Dictyostelium discoidium SS (H.Urushihara)"  
 /dev\_stage="slug"

BASE COUNT 180 a 69 c 71 g 151 t  
 ORIGIN

Query Match 9.58; Score 94.2; DB 106; Length 471;  
 Best Local Similarity 57.48; Pred. No. 1.2e-12;  
 Matches 190; Conservative 0; Mismatches 138; Indels 3; Gaps 1;

468 tcaattgtatcaagaacaatcaatcaagaagcccaagcttggagtcagccagtg 527  
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
 48 TCATCTTTTGTAGTTAAACATCAAGGTTCAAGAAATCCATCTTGGAGG---GAATCTAA 104

528 tataagtagaactagagacgaatctatacagatatgaagaacatttgaagaatat 587  
 105 AATTACAAAGAACCAAGAAAGAGCAATCCGCTAAACTTAATGATATATGAGCAACAATAT 164  
 588 gagtgtgaggttaactaagtgaatttggcaaataccgaagtgatgagctacata 647  
 165 ATCTGTTTCTCCACATTTTGAAGATTAGTCATPAAAAATTCACATGTAGCACTGCTAA 224  
 648 cagagtgagtgatttaaggttttttagcaagaagaatgaaccaccatttgaagaagc 707  
 225 AAGAGGTGGCTATCTTATPCCATTCAAAAGAGGCCAAATGCAAGACCTTTTGAAGATTG 284  
 708 cgaatcaatttgcatttggagagtcagtaacataatgaacccaatagtgccca 767  
 285 CGCATTTTCTTTAAAGTTGCTGAAGTTAGTGGCATTTGTAGATCTGATTCAGAGTCTCA 344  
 768 tateccaaagaacagataaatcaagata 798  
 345 TATATTGAAAGATTACATTAATGTAATTA 375

RESULT 8  
 LOCUS C92940 517 bp mRNA EST 12-JUL-1999  
 DEFINITION C92940 Dictyostelium discoidium SS (H.Urushihara) Dictyostelium  
 accession C92940  
 version C92940.1 GI:3074816  
 keywords EST.  
 source Dictyostelium discoidium.  
 organism Dictyostelium discoidium.  
 reference Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
 authors 1 (bases 1 to 517)  
 title Morio, T., Urushihara, H., Saito, T., Uegawa, Y., Mizuno, H., Yoshida, M.,  
 Yoshino, R., Mitra, B.N., Pl, M., Saito, T., Takemoto, K., Takakawa, H.,  
 Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.  
 journal The Dictyostelium developmental cDNA project: generation and  
 analysis of expressed sequence tags from the first-finger stage of  
 development  
 product = 'Dictyostelium discoidium cDNA project in Japan'  
 location/Qualifiers

FEATURES  
 source  
 1. 517  
 /organism="Dictyostelium discoidium"  
 /strain="AX4"  
 /db\_xref="taxon:44689"  
 /clone\_lib="SSFL124"  
 /dev\_stage="slug"

BASE COUNT 201 a 78 c 74 g 164 t  
 ORIGIN

Query Match 9.58; Score 94.2; DB 175; Length 517;  
 Best Local Similarity 57.48; Pred. No. 1.2e-12;  
 Matches 190; Conservative 0; Mismatches 138; Indels 3; Gaps 1;

468 tcaattgtatcaagaacaatcaatcaagaagcccaagcttggagtcagccagtg 527  
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
 89 TCATCTTTTGTAGTTAAACATCAAGGTTCAAGAAATCCATCTTGGAGG---GAATCTAA 145  
 528 tataagtagaactagagacgaatctatacagatatgaagaagaacatttgaagaatat 587  
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
 146 AATTACAAAGAACCAAGAAAGAGCAATCGCTTAACCTTAATGAATATATGACCAACAATAT 205

[illegible]

Qy	882	aaaactgt	889
Db	667	AGATCTTT	674

RESULT	11
LOCUS	AM010527
DEFINITION	AM010527 708 bp mRNA EST 10-SEP-1999
ACCESSION	U707611 Pine Triplex shoot tip library Pinus taeda cDNA clone
VERSION	U707611, mRNA sequence.
KEYWORDS	AM010527
SOURCE	AM010527.1 GI:5859305
ORGANISM	EST.
	loblolly pine.
	Pinus taeda

REFERENCE	1 (pages 1 to 70b)
AUTHORS	Whetten, R.W., Kinlaw, C.S., Retzel, E. and Sederoff, R.R.
TITLE	The Pine Gene Discovery Project
JOURNAL	Unpublished (1999)
COMMENT	Contact: Ross Whetten

Email: [rosseth@unity.ncsu.edu](mailto:rosseth@unity.ncsu.edu)  
Seq primer: 5' lambda Triplex2 Sequencing Primer

```

source
1. .708
/organism="Pinus taeda"
/db_xref="taxon:3352"
/clone="ST07G11"
/clone_lib="Pine Triplex shoot tip library"
/lab_host="E. coli BM25.8"
/note="Organ: shoot tips; Vector: lambda triplex; Site:1;
Site (A); Site_2: Site (B); Shoot tips (approx. 2 cm from
apex) were collected during the spring, frozen and used
for mRNA isolation. The SMART-PCR method (Clontech) was
used to prepare a library from 1 ug total RNA, using the
lambda triplex vector. Plasmid subclones in pTriplex were
recovered by cre-lox excision in E. coli strain BM25.8 and
sequenced from the 5' end."
BASE COUNT
215 a 137 c 165 g 163 t 28 others
ORIGIN

```

BASE COUNT	215 a	137 c	165 g	163 t	28 others
ORIGIN					
Query Match	9.2%	Score 90.6;	DB 32;	Length 708;	
Best Local Similarity	58.3%	Pred. No. 9,1e-12;			
Matches 201; Conservative	0;	Mismatches 134;	Indels 10;	Gaps 2	

QY	455	aggttaaggttccttcattcttgatcgaagaacaatcaatcaagaaacccaagctcttga	514
Db	128	AGGTCAAGAGCTTCTCATTGCTTATCAACAACAGAGGATCAAGAGGCTTCTCTGTGGC	187
QY	515	agtcceccagatggtatlaag-----tagaactagagaogaatctatacagatatitga	565
Db	188	AAGAGCCCGATGTGGCCGAGATTAAAGCTACCCACGAGATGCTCCGTGCCCACTCA	247
QY	566	agaacactcttggaaagaaatattgagtggtggttaaaactaaagtbaatttggnaataacgc	625
Db	248	GTCGATTTCAGGGAGAAATTTGTTAGTGGCAAGGCCACAGTTTGAGGACCTTGGGCTCGAT	307
QY	626	aaagtbatctcagctcacaatgacagaggttggtgatctttaggtttttagcaaaaggaaca	685
Db	308	ATTCCGACTGTAAATCCCGCTAAGAAAGCGGAGATTTGGTCCATTGGTGCAGGTCAAA	367
QY	686	tgcaacacacactctgaaagaacgcgactcaatttctcagttgggaaa- <i>g</i> tgcaagttaacata	744
Db	368	TGCAAAGCCCTTTTGAAATGCTACATACCTTTTGAATGTCGCAAGATTAAGTATATT	427

Qy 745 attgaaaccaatagltgltccatatcctccaagaacagataa 789  
| | | | | | | | | | | | | | | | | | | | | |  
Db 428 GTGACACTGACAGTGGTGTGCATATCATCTTAAGACAGATGA 4722

RESULT 12	
AV400559	
LOCUS	AV400559 738 bp mRNA EST 05-FEB-2000
DEFINITION	AV400559 Bombyx mori brain Dalzo P0 (just after pupation) Bombyxyxx clone br-1449 T3, mRNA sequence.
ACCESSION	AV400559
VERSION	AV400559.1 GI:6904211
KEYWORDS	EST.
SOURCE	domestic silkworm.
ORGANISM	Bombyx mori

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (pages 1 to 738)	Mita, K., Morimoto, M., Shimada, T., Okano, K. and Maeda, S	Bombyx mori CDNA	Unpublished (2000)	Contact: Mita K

method:uni-directional, sequence direction:sequenced from T3 primer  
(5' -> 3')  
Project:'Silkworm Genome Program in MAFF, and Research for the  
Future Program in JSPS', see 'Silkbase',  
<<http://www.ab.a.u-tokyo.ac.jp/silkbase/>>, for whole ESTdb.

```

FEATURES
source
1..738
Location/Qualifiers
/organism="Bombyx mori"
/strain="Daizo"
/db_xref="taxon:7091"
/clone="br-1449"
/clone_lib="Bombyx mori brain Daizo P0 (just after
pupation)"
/sex="female/male mixed"
/tissue_type="brain"
/dev_stage="P0 (just after pupation)"
255 a 140 c 148 g 195 t
BASE COUNT
ORIGIN
```

BASE COUNT	255 a	140 c	148 g	195 t
ORIGIN				
Query Match	9.1%	Score 90.4	DB 29	Length 738
Best Local Similarity	54.6%	Pred. NO. 1e-11		
Matches 201: Conservative	0	Mismatches 166	Indels 1	Gaps 1

QY	522	agaatggttaagtagtaactagaagacgaatcatcacaatattgaagaacatttgaagaag	581
Db	349	AGAGCATATTACAGTACTAGCTAAGGAGGAAGCTCTTGATATATTACAGAGTATGCCGAA	408
QY	582	aatattgaatgtgtaaggtttaactaaagttaatttggcaaatacggaagtgtattgcagtc	641
Db	409	AATCATTTGACAGAAAGACTTAATTTGAAGAGCTGGCAAGTACATATTTCTGATGTCATTC	468
QY	642	acatgcacgaagtggtgattagggttttttagcaagaagacaatgcaaccacattcga	701
Db	469	AGCAAAACGTATGTGGGATTTGGGTGCTTTTCACAGAAAGTCAAAATGCGAAGAACCATTTGA	528
QY	702	agaagccgcatlcaaatctgcacgttttggagaagtlcagtaacataatgynaaccaatagtvg	761
Db	529	AGAGCTGCATTTCTTCCTTGAAAAATAGGCAATTTAGGCCAACCATTGCACCTGACTCTGG	588
QY	762	tgatcatatctctcaagaagaacaggtataatcaagaataitggagttgattgaanaatlgaa	821
Db	589	CATTTCATTTCTTTTTCAGAACTGCTCTAAAGTTCTTTTTCACAGCTTAACTACTTTAAA	644
QY	822	ataaatgagaagaagtgtgatagatttggtaaccaaaaaagcgaatggtctcacaagaatgc	881



Db	648	AAGGTAAACAAATATTCGA	CTTATCTTACCCACATATCG	CAATACATCAAGTAT	707
Qy	882	aaactgt	889		
Db	708	AGATCTTT	715		

RESULT 13

LOCUS	AU003558	748 bp	mRNA	EST	19-JAN-1999
DEFINITION	AU003558	Bombyx mori	p50(Dalzo)	Bombyx mori	clone ws00315.
ACCESSION	AU003558		mRNA sequence.		
VERSION	AU003558				
KEYWORDS	AU003558.1	GI:4160696			
ORGANISM	EST.				
SOURCE	domestic silkworm.				
	Bombyx mori				

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (pages 1 to 74)  
Mita, K., Moriyama, M., Shimada, T., Okano, K. and Maeda, S  
Establishment of cDNA database of Bombyx mori  
Unpublished (1999)  
Contact: Mita K

## RESULT 14

LOCUS	660 bp	mRNA	EST	15-JAN-1999
AU000274				
DEFINITION	Bombyx mori	p50(Daizo)	Bombyx mori	CDNA clone e40334, mRNA sequence.

National Institute of Radiological Sciences  
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan  
Email: kmitsue@nirs.go.jp  
PROJECT = 'CREST project by JST',  
Location/Qualifiers

**RESULT 15**

LOCUS	AU003857	722 bp	mrna	EST	19-JAN-1999
DEFINITION	AU003857	Bombyx mori	p50(Dalzo)	Bombyx mori	cdna clone w500674
					mrna sequence.

QY	882	aaaacgt	889
Db	688	AGATCTTT	695

ACCESSION	AU003857
VERSION	AU003857.1
KEYWORDS	GI:4161228
SOURCE	EST.
ORGANISM	domestic silkworm. Bombyx mori
	Eukaryote; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia

```
REFERENCE      ; Bombycoidea; Bombycidae; Bombyx.
AUTHORS        Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.
TITLE          Establishment of cDNA database of Bombyx mori
JOURNAL        Unpublished (1999)
COMMENT        Contact: Mita K
                Genome Research Group
                National Institute of Radiological Sciences
                Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
                Email: kmitta@nirs.go.jp
                PROJECT = 'CREST project by JSR'.

FEATURES
  source
    1..722
    /organism="Bombyx mori"
    /strain="p50(Daizo)"
    /db_xref="taxon:7091"
    /clone="w80674"
    /clone_1bp="Bombyx mori p50(Daizo)"

BASE COUNT     246 a      141 c      145 g      190 t
ORIGIN

Query Match
Best Local Similarity 9.1%; Score 89.8; DB 106; Length 722;
Matches 163; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 522 agatgylataagtagaactagagacgaatctatacagataltgaagaacattggaag 581
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 318 AGACGATTTTACAGCTACTAGAGGAGGAGCTCTGATATATTACAGAGTATCGCCCTAA 377

QY 582 aatattgagtgtaggttaactaagtgaattggcaaataccgaagaattgcaagctc 641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 378 AATCATTCACAGAGAGAGCTAAATTGAAGAGCTGGCAGTACATATCTGATTGTTCAATC 437

QY 642 acatgacagagtgtaggttaagggttttagcaagaagacaatgcaaccacattcga 701
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 AGCAAAACGTCATGGGATTTGGGTCGTTTCAAGAAAGGTCAAATGCGAAGAACCATTTGA 497

QY 702 agaagccgcatcattgcatgtltggaagtcagtaacataattgaaaccaatagtcg 761
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 498 AGACGTGCGATTTTCCCTGGAATAATAGGCAATTTGAGCCACAGTTCACACTGACTCTGG 557

QY 762 tgtccatatctccaagaacagagataaatacaagatatltgaggt 806
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 558 CATTCATATCATCTTAGAACTGCCTAAAGATTGTTTTCACAGTT 602
```

Search completed: June 8, 2001, 05:10:35  
Job time: 4325 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2001, 09:37:50 ; Search time 22.35 Seconds

(Without alignments)  
928.224 Million cell updates/sec

Title: US-09-507-242-2

Perfect score: 932  
Sequence: 1 MASTSTGLPMTIRVSRSH.....GEVSNITETNSGVHLQRTG 177

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP\_TREMBL\_15:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organella:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rhodent:\*  
13: sp\_unclassified:\*  
14: sp\_vertebrate:\*  
15: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	448.5	48.1	176	3 042735	042735 emericella
2	419.5	45.0	182	3 060045	060045 neurospora
3	409	43.9	175	3 074448	074448 schizosacch
4	373.5	40.1	159	13 0919K6	0919K6 xenopus lae
5	361.5	38.8	165	11 090UR7	090UR7 mus sp. pin
6	329.5	35.4	168	5 09N492	09N492 caenorhabdi
7	325	34.9	166	5 09VRH1	09VRH1 drosophila
8	305.5	32.8	118	10 09LEK8	09LEK8 digitalis l
9	303.5	32.6	119	10 09SL42	09SL42 arabidopsis
10	271.5	29.1	243	5 P90527	P90527 dictyosteli
11	211	22.6	69	10 042334	042334 arabidopsis
12	166.5	17.9	333	2 09KDN4	09KDN4 bacillus ha
13	157	16.8	92	1 074049	074049 cenarchaeum
14	142	15.2	221	10 023727	023727 arabidopsis
15	122	13.1	347	2 09JSP0	09JSP0 neisseria m
16	122	13.1	348	2 09K186	09K186 neisseria m
17	121.5	12.8	299	2 09ZM07	09ZM07 helicobacte
18	119.5	12.8	92	2 09KKZ7	09KKZ7 vibrio chol
19	118.5	12.7	655	2 09PE37	09PE37 xylella fas

20	117.5	12.6	282	2 09ZCX6	09ZCX6 rickettsia
21	116.5	12.5	336	2 051135	051135 borrelia bu
22	116.5	12.5	619	2 09KOR0	09KOR0 vibrio chol
23	109	11.7	93	2 09L6S3	09L6S3 salmonella
24	108	11.6	431	2 09KUS0	09KUS0 vibrio chol
25	107	11.5	264	2 052073	052073 enterobacte
26	106	11.4	250	2 054047	054047 pseudomonas
27	103	11.1	622	2 031248	031248 acinetobact
28	102	10.9	130	5 09VB04	09VB04 drosophila
29	102	10.9	131	4 09V237	09V237 homo sapien
30	97	10.4	577	11 09JLV1	09JLV1 mus musculu
31	95	10.2	126	5 09NAR9	09NAR9 caenorhabdi
32	94	10.1	577	11 09JVC7	09JVC7 mus musculu
33	87.5	9.4	464	2 09PFA0	09PFA0 xylella fas
34	87.5	9.4	1246	5 09VXK9	09VXK9 drosophila
35	87	9.3	376	11 061048	061048 mus musculu
36	85.5	9.2	879	2 09ZJN8	09ZJN8 helicobacte
37	85	9.1	1585	4 043432	043432 homo sapien
38	84.5	9.1	709	5 021376	021376 caenorhabdi
39	84.5	9.1	877	2 026042	026042 helicobacte
40	84	9.0	150	2 034682	034682 bacillus su
41	84	9.0	519	10 P93833	P93833 arabidopsis
42	84	9.0	720	10 09SQM6	09SQM6 arabidopsis
43	83.5	9.0	911	3 006704	006704 saccharomyc
44	83	8.9	566	3 09P766	09P766 schizosacch
45	82.5	8.9	61	2 045680	045680 bacillus su

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	176 AA.
042735	042735			
AC	042735:			
DT	01-JUN-1998 (TREMBLrel. 06, Created)			
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
DE	PEPTIDYL-PROLYL CIS/TRANS ISOMERASE.			
GN	PINA.			
OS	Emericella nidulans (Aspergillus nidulans).			
OC	Eukaryota; Fungi; Ascomycota; Eurotiiales; Trichocomaceae; Emericella.			
OX	NCBI_TaxID=5072;			
ON	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-R153:			
RX	MEDLINE=98151356; PubMed=9482729;			
RA	Crenshaw D.G., Yang J., Means A.R., Kornbluth S.;			
RT	"The mitotic peptidyl-prolyl isomerase, Pin1, interacts with Cdc25 and			
RT	Pin1."			
RL	EMBL J. 17:1315-1327(1998).			
DR	EMBL; AF035768; AAC49984.1; -			
DR	HSSP; Q13526; PIN.			
DR	INTERPRO; IPR000297; -			
DR	INTERPRO; IPR001202; -			
DR	INTERPRO; IPR002349; -			
DR	PFAM; PF00397; WW; 1.			
DR	PFAM; PF00639; Rotamase; 1.			
DR	PRINTS; PR00403; WMDOMAIN.			
DR	PROSITE; PS01159; WW_DOMAIN_1; UNKNOWN_1.			
DR	PROSITE; PS00020; WW_DOMAIN_2; 1.			
DR	PRODOM; PD002510; -; 1.			
DR	Isomerase.			
SC	SEQUENCE 176 AA; 20038 MW; 1D1E8376239E1309 CRC64;			
Query Match	48.1%; Score 448.5; DB 3; Length 176;			
Best Local Similarity	47.4%; Pred. No. 3.4e-37;			
Matches	82; Conservative 40; Mismatches 48; Indels 3; Gaps 2;			
QY	5 STGLPMTIRVSRSHNKKEYFLNOSTNESSWDPPYGTDEVLNAYIAFKNNG--YKPLV 62			
DB	3 NTGLPAGWEVHRSHNSKNLPYFNPATRESRWEPPADDTMETLKMYYMATYHSGAATYHEAP 62			

```

OY 63 NEDGQVAVSHLLIKNNQSRKPKSKSPDGIISRTDESIQILKKHLRIISGEVKLSLAN 122
DB 63 SOEGKIRCSHLVYKHDRSRPSSWREAE-ITRTKEARELLRHQRIRMGELRLDLM 121
OY 123 TSDSCSHRGGDLGFFSGKQMPPEFEAFNLHGEVSNIIETNSGVHILOR 175
DB 122 SEDSCSARRKGDGLGFFGEGMOKEFEFAFALOPQVSDIVSGSGHLIER 174

RESULT 2
060045 PRELIMINARY; PRT; 182 AA.
AC 060045;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, last annotation update)
DE PEPTIDYLPROLYL ISOMERASE (EC 5.2.1.8).
GN SSP1.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Kops O., Tropeschug M.;
RT Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ006023; CAA06818.1; -.
DR HSP; Q13526; IPIN.
DR INTERPRO: IPR000297; -.
DR INTERPRO: IPR001202; -.
DR INTERPRO: IPR002349; -.
DR PFAM; PF00397; WW; 1.
DR PFAM; PF00639; Rotamase; 1.
DR PRINTS; PR00403; WMDOMAIN.
DR PROSITE; PS01159; WW_DOMAIN_1; UNKNOWN_1.
DR PROSITE; PS50020; WW_DOMAIN_2; 1.
DR PROMOD; PD002510; -. 1.
KW Isomerase.
SQ SEQUENCE 182 AA; 20673 MW; AC8D0497DEFF7339 CRC64;

Query Match
Best Local Similarity 44.3%; Score 419.5; DB 3; Length 182;
Matches 82; Conservative 40; Mismatches 48; Indels 15; Gaps 4;

OY 1 MAST-STGLPWNMTIRVSRSHNKEYFLNOSTNESSWDPPYGTDEKVLNAYIAFKNNGYK 59
DB 1 MSNTIETGLPEDEMEVYHSGSKMLPYFNATKTSRKEPPSGTDVDKLIYMAKY----HS 56
OY 60 PLVNE-----DQGVAVSHLLIKNNQSRKPKSKSPDGIISRTDESIQILKKHLERI 110
DB 57 PYSQOQOQOQOQOPOGKICAHLLVKNQSRPSSWRESE-ITRTKEARELLRHQRIRMGELRLDLM 115
OY 111 LSGEVLISLANTESDSCSHDRGGDLGFFSGKQMPPEFEAFNLHGEVSNIIETNSGV 170
DB 116 KGSISLGLALATESDSCSARRKRGDLGTRGDMKEFEFAFALPKPGELISDIVTASGL 175
OY 171 HILQR 175
DB 176 HLIER 180

RESULT 3
074448 PRELIMINARY; PRT; 175 AA.
AC 074448;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, last annotation update)
DE PROBABLE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE (EC 5.2.1.8).
GN SPCC1604.03.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;

```

```

OC Schizosaccharomycetales; Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-972;
RA Purnelle B., Goffeau A., Wood V., Rajandream M.A., Barrell B.G.;
RT Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC PEPTIDE BONDS IN OLIGOPEPTIDES.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE PPIC/PARVULIN FAMILY OF ROTAMASES.
CC -1- SIMILARITY: CONTAINS 1 WW DOMAIN.
DR EMBL; AL031535; CAA20742.1; -.
DR HSP; Q13526; IPIN.
DR INTERPRO: IPR000297; -.
DR INTERPRO: IPR001202; -.
DR PFAM; PF00397; WW; 1.
DR PFAM; PF00639; Rotamase; 1.
DR PROSITE; PS50020; WW_DOMAIN_2; 1.
DR PROMOD; PD002510; -. 1.
DR Isomerase; Rotamase; Nuclear protein.
KW DOMAIN 4 38 WW_DOMAIN.
FT DOMAIN 64 175 PPIC-LIKE.
SQ SEQUENCE 175 AA; 19773 MW; ABA637835471BD25 CRC64;

Query Match
Best Local Similarity 43.9%; Score 409; DB 3; Length 175;
Matches 80; Conservative 37; Mismatches 53; Indels 4; Gaps 4;

OY 4 TSTGLPWNMTIRVSRSHNKEYFLNOSTNESSWDPPYGTDEKVLNAYIAFKNNGYKPL-V 62
DB 2 SNTGLPKPWIVKISRSRNPYFENFTHESLWEPATDMAALKRIANELLQESVYPTREA 61
OY 63 NEDGQVAVSHLLIKNNQSRKPKSKSPDGIISRTDESIQILKKHLRIISGEVKLSLA 121
DB 62 SNSPKIRASHLLVKNHESRSPSSWKE-EHITRSKEBA-RKLEHETQLLKSGSVSMHDA 119
OY 122 NTESDSCSHDRGGDLGFFSGKQMPPEFEAFNLHGEVSNIIETNSGVHILOR 175
DB 120 MKESDSCSARRGELGFEFRDEMOKEFEFAFALPKPGELISGVETSSGHHIOR 173

RESULT 4
0919K6 PRELIMINARY; PRT; 159 AA.
AC 0919K6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, last annotation update)
DE PROLYL ISOMERASE PIN1.
GN PIN1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8335;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20165035; PubMed=10698738;
RA Winkler K.E., Swenson K.I., Kornbluth S., Means A.R.;
RT "Requirement of the prolyl isomerase pin1 for the replication
RT checkpoint."
RT Science 287:1644-1647 (2000) .
DR EMBL; AF239760; AAF43897.1; -.
KW Isomerase.
SQ SEQUENCE 159 AA; 17671 MW; 4E54F66F16C9DF1C CRC64;

Query Match
Best Local Similarity 44.4%; Score 373.5; DB 13; Length 159;
Matches 75; Conservative 32; Mismatches 45; Indels 17; Gaps 3;

```

Query Match 38.8%; Score 361.5; DB 11; Length 165;  
Best Local Similarity 42.6%; Pred. No. 1.5e-24;  
Matches 72; Conservative 35; Mismatches 51; Indels 11; Gaps 3;

QY 8 LPPNNTIVSRSHNKKEYFLNOSTNESSMDPPYGTDKVELNAYIAFKNNGYKPLVNEGQ 67  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10095;

QY 7 LPPGWEKRRSRSGRYVFNHTNATASQWERP-----IAGCKNGQGEF-----GK 50  
OC NCBI\_TaxID=6239;

QY 68 VVSHLLIKNNOSRKRKPKSMKSPDGISRTDESIQILKHLERLISGEVKSLELNTESDC 127  
RC STRAIN-BRISTOL N2;  
MEDLINE=99069613; PubMed=9851916;  
None.

QY 128 SSHRDGDLGFFSKGOMQPFEEAFLNHYGVSNIIETNSGVHILORT 176  
RT Investigating Biology. The C. elegans Sequencing Consortium.  
Science 282:2012-2018(1998).

QY 110 SSAAKAGDGLGFFSKGOMQPFEDASFLALPBGMSGPVFTDGIHILRT 158  
RL SRAIN-BRISTOL N2;  
Waterston R.;  
Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
EMBL: AC024746; MAF60409.1; - OFC837780579C070 CRC64;  
SEQUENCE 168 AA; 19176 MW; 188E95F009176B1F CRC64;

RESULT 5  
Q9OUR7 PRELIMINARY; PRT; 165 AA.  
ID Q9OUR7;  
AC Q9OUR7;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
DE PINT.  
GN MPIN.  
OS Mus sp.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10095;

RN [1]  
RA SEQUENCE FROM N.A.  
FUJIMORI F., Takahashi K., Uchida C., Uchida T.;  
"Mice lacking Pnt1 develop normally, but are defective in entering  
cell cycle from G0 arrest."  
BIOchem. Biophys. Res. Commun. 265:658-663(1999).

RL EMBL: AB009692; BAA87038.1; -  
DR EMBL: AB009691; BAA87037.1; -  
DR HSSP: Q13526; 1PIN.  
DR INTERPRO: IPR000297; -  
DR INTERPRO: IPR001202; -  
DR INTERPRO: IPR002349; -  
DR PFAM: PF00397; WW; 1.  
DR PFAM: PF00639; Rotamase; 1.  
DR PRINTS: PRO0403; WMDOMAIN.  
DR PROSITE: PS01096; PRIC\_PRINSE; 1.  
DR PROSITE: PS01159; WW\_DOMAIN\_1; 1.  
DR PROSITE: PS50020; WW\_DOMAIN\_2; 1.  
SQ SEQUENCE 165 AA; 18370 MW; 188E95F009176B1F CRC64;

Query Match 35.4%; Score 329.5; DB 5; Length 168;  
Best Local Similarity 39.2%; Pred. No. 9.8e-22;  
Matches 73; Conservative 32; Mismatches 46; Indels 35; Gaps 5;

QY 4 TSTGLPPNNTIVSRSHNKKEYFLNOSTNESSMDPP-----YGTKEVLNAYIAFKNNGYK 59  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10095;

QY 2 SDNSLPAGMEKRSRSGRYVFNHTNATASQWERP-----IAGCKNGQGEF-----GK 50  
OC NCBI\_TaxID=6239;

QY 60 PLVNEGQGVRSVSHLLIKNNOSRKRKPKSMKSPDGISRTDESIQILKHLERLISG 113  
RC STRAIN-BRISTOL N2;  
MEDLINE=99069613; PubMed=9851916;  
None.

QY 114 ----EVLKSELNTESDCSSHDGDLGFFSKGOMQPFEEAFLNHYGVSNIIETNSG 169  
RT Investigating Biology. The C. elegans Sequencing Consortium.  
Science 282:2012-2018(1998).

QY 101 ASNIECKFREILAKFSDSCSAKRGDLGPFEROMQPFEDASFLALPBGMSGPVFTDGIHILRT 160  
RL SRAIN-BRISTOL N2;  
Waterston R.;  
Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
EMBL: AC024746; MAF60409.1; - OFC837780579C070 CRC64;  
SEQUENCE 168 AA; 19176 MW; 188E95F009176B1F CRC64;

RESULT 7  
Q9VRH1 PRELIMINARY; PRT; 166 AA.  
ID Q9VRH1;  
AC Q9VRH1;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
DE DOD PROTEIN.  
GN DOD.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_TaxID=7227;

RN [1]  
RA SEQUENCE FROM N.A.  
MEDLINE=20196006; PubMed=10731132;  
Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Amann M.D., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Mortman J.R., Lander M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,  
Aburil J.F., Agbayani A., An H.-J., Andrews-Planck C., Baldwin D.,  
Bailly R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.T., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
Borkova D., Borkova M.R., Bouck J., Broksstein P., Brothier P.,  
Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dudin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Idegwan C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Liao P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matlet B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nussken D.R., Paclet J.M.,  
 RA Palazolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA She B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissensbach J.,  
 RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL, AE003568; AAF50829.1; -.  
 DR HSSP, Q13526; IPTN.  
 DR FLYBASE; FBgn0015379; dod.  
 DR INTERPRO; IPR000297; -.  
 DR INTERPRO; IPR001202; -.  
 DR PFAM; PF00397; MW; 1.  
 DR PFAM; PF00639; Rotamase; 1.  
 DR PROSITE; PS01096; PPIQ\_PPIASE; 1.  
 DR PROSITE; PS01159; MW\_DOMAIN.1; 1.  
 DR PROSITE; PS50020; MW\_DOMAIN.2; 1.  
 SQ SEQUENCE 166 AA; 18376 MW; 3B4306FA930E7259 CRC64;

Query Match 34.9%; Score 325; DB 5; Length 166;  
 Best Local Similarity 41.7%; Pred. No. 2.4e-21;  
 Matches 70; Conservative 30; Mismatches 58; Indels 10; Gaps 3;  
 QY 8 LPPNNTIVSRSHNKEFLPNOSTNSSMDPPYGTDEKVLNATIAFKNGNGYPLVNEGQ 67  
 DB 7 LPDGKEKTRSRSTGMSYLYMNTKESQMDP---AKKAGGSGAGGDAPE 57  
 QY 68 VVSHLLIKNNOSRKRKSPKSPGDISRTRDESIOILKKHLERILSGEYKLSFLANTESDC 127  
 DB 58 VHCILLIKKHKSRKSPKSPKREAN-ITRTKEAQLLEIYRKTIVOQEAFTDELANSYSDC 116  
 QY 128 SSHDRGDLGFFSGKQMPPEEAAFNHLVGEVSNIIETNSGVHILORT 175  
 DB 117 SSARRGGDLGKFRGQMOAFAEDAFKLVNQLSGIVSDSLHILR 164

RESULT 8  
 Q9LEK8 PRELIMINARY; PRT; 118 AA.  
 AC Q9LEK8;  
 DT 01-OCT-2000 (TREMBlrel. 15 Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE.  
 GN PARY12.8.  
 OS *Digitalis lanata* (Foxglove).  
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 CC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
 CC Lamiales; Scrophulariaceae; Digitalis.  
 OX NCBI\_TaxID=49450;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HRH;  
 RA Metzner M., Ruecknagel P., Kuelbert G., Luckner M.;

RT "Isolation and characterization of a new parvulin like peptidyl-prolyl  
 RT cis/trans isomerase from *Digitalis lanata*.";  
 RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AJ133755; CAB94994.1; -.  
 KW Isomerase.  
 SQ SEQUENCE 118 AA; 12834 MW; 131B74FB4AC3F229 CRC64;

Query Match 32.8%; Score 305.5; DB 10; Length 118;  
 Best Local Similarity 52.6%; Pred. No. 8.2e-20;  
 Matches 60; Conservative 21; Mismatches 30; Indels 3; Gaps 1;  
 QY 67 QVRVSHLLIKNNOSRKRKSPKSPGDISRTRDESIOILKKHLERILSGEYKLSFLANT 123  
 DB 5 KVRASHLLIKKHKSRKSPKSPKDPDLSLATRRDDAVSOLSLROELLSDPASFDLASR 64  
 QY 124 ESDCSSHDRGDLGFFSGKQMPPEEAAFNHLVGEVSNIIETNSGVHILORTG 177  
 DB 65 HSHCSSARRGGDLGFFSGKQMPPEEAAFNHLVGEVSNIIETNSGVHILRTG 118

RESULT 9  
 Q9SLA2 PRELIMINARY; PRT; 119 AA.  
 AC Q9SLA2;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE PUTATIVE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE.  
 GN AT2G18040.  
 OS *Arabidopsis thaliana* (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 CC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 CC Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, COLUMBIA;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffatt K.S.,  
 RA Cronin L.A., Shen M., Vanden S.E., Umayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carreira A.D., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome II of *Arabidopsis thaliana*.";  
 RL Nature 402:761-768(1999).  
 DR EMBL, AC006201; AAD20122.1; -.  
 DR HSSP, Q13526; IPTN.  
 DR INTERPRO; IPR000297; -.  
 DR PFAM; PF00639; Rotamase; 1.  
 DR PROSITE; PS01096; PPIQ\_PPIASE; 1.  
 KW Isomerase.  
 SQ SEQUENCE 119 AA; 13015 MW; E926CB566E76A0A3 CRC64;

Query Match 32.6%; Score 303.5; DB 10; Length 119;  
 Best Local Similarity 51.3%; Pred. No. 1.2e-19;  
 Matches 58; Conservative 24; Mismatches 28; Indels 3; Gaps 1;  
 QY 67 QVRVSHLLIKNNOSRKRKSPKSPGDISRTRDESIOILKKHLERILSGEYKLSFLANT 123  
 DB 6 QVKASHLLIKKHKSRKSPKSPKDPDLSLATRRDDAVSOLSLROELLSDPASFDLASR 65  
 QY 124 ESDCSSHDRGDLGFFSGKQMPPEEAAFNHLVGEVSNIIETNSGVHILORT 176  
 DB 66 VSDCSSARRGGDLGFFSGKQMPPEEATYALKVDISIDYVDTSGVHILKRT 118  
 RESULT 10  
 P90527 PRELIMINARY; PRT; 243 AA.  
 AC P90527;  
 AC P90527;

DT 01-MAY-1997 (Tremblrel. 03, Created)  
DT 01-MAY-1997 (Tremblrel. 03, last sequence update)  
DT 01-JUN-2000 (Tremblrel. 14, last annotation update)  
GN PINA.  
OS Dictyostellum discoideum (Slime mold).  
OC Eukaryota; Dictyostelida; Dictyostellum.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AX4;  
RA Loomis W.F., Iranfar N.;  
RL Submitted (Nov-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U78757; AAB36960.1; -.  
DR HSSP; Q13526; IPIN.  
DR INTERPRO; IPR000253; -.  
DR INTERPRO; IPR000297; -.  
DR PFAM; PF00498; FHA; 1.  
DR PFAM; PF00639; Rotamase; 1.  
DR PROSITE; PS50006; FHA\_DOMAIN; 1.  
DR NON\_TER 1  
SQ SEQUENCE 243 AA; 26794 MW; 1596001BB784500A CRC64;

Query Match 29.1%; Score 271.5; DB 5; Length 243;  
Best Local Similarity 37.4%; Pred. No. 2e-16;  
Matches 65; Conservative 29; Mismatches 57; Indels 23; Gaps 4;

OY 4 TSGTGLPPNNTIRSRSHNKEYFL--NOSTNESSWDPPYGTGDEKVLNAYIKFKNNGYKPL 61  
DB 89 TPTTVENETTFKFG-SSSKHFLIKGTNTNPSS-----SSSSSSSS 128  
OY 62 VNEDGVRSHLLIKNNOSRPPKSPDGISRTDESIOILKKHLRIISGEVKLSLA 121  
DB 129 SSSEPKVYTCRHLLVKNQSGSNPSSMRE-SKITTKERAIAKLMEYRATITSGSNTPEDLA 167  
OY 122 NTESDSSHDGGLGFFSKGQMPPEEAFLHVGESNIETNSGVHILOR 175  
DB 188 HKNSDCSSAKRGGLDLPFRKGQMRPEFDCAFSLKGVESGIYDTSGVHIER 241

RESULT 11  
O42334 PRELIMINARY; PRT; 69 AA.  
AC O42334;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)  
DT 01-JUN-2000 (Tremblrel. 14, last annotation update)  
DE PEPTIDYL PROLYL CIS-TRANS ISOMERASE (FRAGMENT).  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Tridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-DRY SEEDS OF A. THALIANA ECOTYPE COLUMBIA;  
RA Cooke R., Laudie M., Raynal M., Delseny M.;  
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; P13319; CAA23077.1; -.  
DR HSSP; Q13526; IPIN.  
DR MENDEL; 12876; Arabid.2076;12876.  
DR INTERPRO; IPR000297; -.  
DR PFAM; PF00639; Rotamase; 1.  
DR PROSITE; PS01096; PPICT\_PPIASE; 1.  
KW Isomerase.  
FT NON\_TER 1  
SQ SEQUENCE 69 AA; 7333 MW; 03C2DF735ED6F253 CRC64;

Query Match 22.6%; Score 211; DB 10; Length 69;  
Best Local Similarity 58.2%; Pred. No. 8.9e-12;  
Matches 39; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

OY 110 ILSGEVLSLANTESDSSHDGGLGFFSKGQMPPEEAFLHVGESNIETNSG 169  
DB 2 IVSGKANFEVATRVSDCSSAKRGGDLGSRGQMPPEEAFLYALKVGDIDVIDDSG 61  
OY 170 VHIILQRT 176  
DB 62 VHIILKRT 68

RESULT 12  
O9KDN4 PRELIMINARY; PRT; 333 AA.  
AC O9KDN4;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, last annotation update)  
DE PROTEIN SECRETION (POST-TRANSLATION CHAPERONIN).  
GN PRSA.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C-125 / JCM 9153;  
RA Takami H., Nakasone K., Takaki Y.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF001511; BAB04896.1; -.  
SQ SEQUENCE 333 AA; 37266 MW; A0B77383BC1AD61E CRC64;

Query Match 17.9%; Score 166.5; DB 2; Length 333;  
Best Local Similarity 30.1%; Pred. No. 5.5e-07;  
Matches 50; Conservative 25; Mismatches 40; Indels 51; Gaps 7;

OY 21 NKEYFLNOSTNESSWDPPYGTGK-----EVLNAYIAKFRNNGYKPLVNEGG--OY 68  
DB 114 NKEEFVNEYIL-----PPLVLEKIAVGEVETDEKQAYRDE-----NRDSLIEV 158  
OY 69 RVSHLLIKNNOSRPPKSPDGISRTDESIOILKKHLRIISGEVKLSLANTES-DC 127  
DB 159 EASHILVEDET-----AEVLDRLRAGD-DFEALASEVSYDP 195  
OY 128 SSHDRGDLGFFSKGQMPPEEAFLHVGESNIETNSGVHIL 173  
DB 196 SAANNODLGFFSKGQMPPEEAFLMEIDEVSEPEVSTGYHIT 241

RESULT 13  
O74049 PRELIMINARY; PRT; 92 AA.  
AC O74049;  
DT 01-NOV-1998 (Tremblrel. 08, Created)  
DT 01-NOV-1998 (Tremblrel. 08, last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, last annotation update)  
DE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE.  
GN PINA.  
OS Cenarchaeum symbiosum.  
OC Archaea; Crenarchaeota; Cenarchaeum.  
OX NCBI\_TaxID=46770;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A;  
RX MEDLINE=98422450; PubMed=9748430;  
RA Schleper C., Delong E.F., Preston C.M., Feldman R.A., Wu K.Y.,  
RA Swanson R.V.;  
RT "Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon Cenarchaeum symbiosum.";  
DR J. Bacteriol. 180:5003-5009(1998).  
DR EMBL; AF083071; AAC62692.1; -.  
DR HSSP; Q13526; IPIN.  
DR INTERPRO; IPR000297; -.

```
DR PFAM: PF00639; Rotamase; 1.
DR PRODOM: PD002510; -; 1.
KW Isomerase.
SQ SEQUENCE 92 AA; 10067 MW; AADAB5FE60C518FE CRC64;

Query Match
Best Local Similarity 33.9%; Score 157; DB 1; Length 92;
Matches 38; Conservative 22; Mismatches 28; Indels 24; Gaps 4;

OY 67 OVRVSHLLIKNNQSRKPKSWKSPDGISRTDRDSIOILKKHLERILSGEYKLSLANTES-125
DB 4 KIKCSHLLVKK-----QGEALAV---QERLKAGE-KFGKIAKELSI 40

OY 126 DCSHSDRGDGLGFFSKGQMPPEEAFAFNLHAGEVSNIIETNSGVHILQRTG 177
DB 41 DGSAAKRDSGLGFGKMKVKKPFEDAFAFLQYGEVSEPVKSEFGVHVIRLIG 92

RESULT 14
O23727 PRELIMINARY; PRT; 221 AA.
AC 023727;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Magnoliophyta; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Brassicales; eudicotyledons; core eudicots; Rosidae; eurosids II;
OX NCBI_TaxID=3702;
RN [1]
RP STRAIN=T-DNA TAGGED LINE SK3-1; TISSUE=LEAVES;
RA Babyichuk E., Funangthong M., Inze D., van Montagu M., Kushnir S.;
RU Submitted (FEB-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: Z86095; CAB06699.1; -.
DR HSSP: Q13526; IPIN.
DR INTERPRO: IPR000297; -.
DR INTERPRO: IPR001763; -.
DR PFAM: PF00581; Rhodanese; 1.
DR PFAM: PF00639; Rotamase; 1.
DR PRODOM: PD002510; -; 1.
KW Isomerase.
SQ SEQUENCE 221 AA; 24744 MW; 491BE5CACA5E6117 CRC64;

Query Match
Best Local Similarity 15.2%; Score 142; DB 10; Length 221;
Matches 36; Conservative 17; Mismatches 33; Indels 22; Gaps 3;

OY 67 OVRVSHLLIKNNQSRKPKSWKSPDGISRTDRDSIOILKKHLERILSGEYKLSLANTESD 126
DB 17 EILVQHLLVKNK-----DVELFAELQKKFLDGE-EMSDLALEYSI 55

OY 127 CSSHSDRGDGLGFFSKGQMPPEEAFAFNLHAGEVSNIIETNSGVHILQ 174
DB 56 CPSKRDGILGVMKLGQVPEEEAFAFKAELNQVVR-CRTQGLHILQ 102

RESULT 15
O9JSP0 PRELIMINARY; PRT; 347 AA.
AC O9JSP0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PUTATIVE ROTAMASE.
GN NMA2206.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491".
RL Nature 404:502-506(2000).
DR EMBL: AL162758; CAB85417.1; -.
SQ SEQUENCE 347 AA; 37126 MW; A387AD675F86FA51 CRC64;

Query Match
Best Local Similarity 13.1%; Score 122; DB 2; Length 347;
Matches 35; Conservative 14; Mismatches 41; Indels 22; Gaps 2;

OY 67 OVRVSHLLIKNNQSRKPKSWKSPDGISRTDRDSIOILKKHLERILSGEYK----LSELAN 122
DB 207 QYRAQHILIKAD-----SENAVGAESTRKITYGEARSGTDFSSILAR 248

OY 123 TESDSSHDGDLGFFSKGQMPPEEAFAFNLHAGEVSNIIETNSGVHILQ 174
DB 249 QVSQDASAGNGDGLGFWADGVVWPAFEEAVHALKPGVGVAPVPTQFGWHIIR 300
```

Search completed: June 8, 2001, 09:38:38  
Job time: 48 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2001, 05:10:35 ; Search time 1747.82 Seconds  
(without alignments)  
99.966 Million cell updates/sec

Title: US-09-507-242-3

Perfect score: 20  
Sequence: 1 ccagatggtataagtagaac 20

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 9623517 segs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_esthum29:\*  
64: em\_esthum30:\*  
65: em\_esthum31:\*  
66: em\_esthum32:\*  
67: em\_esthum33:\*  
68: em\_esthum34:\*  
69: em\_esthum35:\*  
70: em\_esthum36:\*  
71: em\_esthum37:\*  
72: em\_esthum38:\*  
73: em\_esthum39:\*  
74: em\_esthum40:\*  
75: em\_esthum41:\*  
76: em\_esthum42:\*  
77: em\_esthum43:\*  
78: em\_esthum44:\*  
79: em\_esthum45:\*  
80: em\_esthum46:\*  
81: em\_esthum47:\*  
82: em\_esthum48:\*  
83: em\_esthum49:\*  
84: em\_esthum50:\*  
85: em\_esthum51:\*  
86: em\_esthum52:\*  
87: em\_esthum53:\*  
88: em\_esthum54:\*  
89: em\_esthum55:\*  
90: em\_esthum56:\*  
91: em\_esthum57:\*  
92: em\_esthum58:\*  
93: em\_esthum59:\*  
94: em\_esthum60:\*  
95: em\_esthum61:\*  
96: em\_esthum62:\*  
97: em\_esthum63:\*  
98: em\_esthum64:\*  
99: em\_esthum65:\*  
100: em\_esthum66:\*  
101: em\_esthum67:\*  
102: em\_esthum68:\*  
103: em\_esthum69:\*  
104: em\_esthum70:\*  
105: em\_esthum71:\*  
106: em\_esthum72:\*  
107: em\_esthum73:\*  
108: em\_esthum74:\*  
109: em\_esthum75:\*  
110: em\_esthum76:\*  
111: em\_esthum77:\*  
112: em\_esthum78:\*  
113: em\_esthum79:\*  
114: em\_esthum80:\*  
115: em\_esthum81:\*  
116: em\_esthum82:\*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

```

190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	16.8	84.0	225	127	BB208038
2	16.8	84.0	611	31	AV655972
3	16.8	84.0	630	214	AO980511
C 4	16.8	84.0	710	16	AI132934
5	16.8	84.0	846	149	BF540037
C 6	16.8	84.0	894	151	BF692672
7	16.4	82.0	188	139	BE807035
C 8	16.4	82.0	213	16	AI120952
9	16.4	82.0	222	11	AA717739
10	16.4	82.0	242	7	AA458264
11	16.4	82.0	250	149	BF542933
C 12	16.4	82.0	271	152	BF812297
13	16.4	82.0	281	152	F06213
14	16.4	82.0	291	152	F06677
15	16.4	82.0	298	113	AW313588
16	16.4	82.0	308	152	F06211
17	16.4	82.0	320	155	R22624
18	16.4	82.0	331	113	AW313587



Tel: 86-21-50801919(ex.45)  
Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn  
This clone is available at CHGC in Shanghai.

FEATURES  
source  
Location/Qualifiers

1. 611  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="GICEMD12"  
/clone\_1lb="GIC"  
/tissue\_type="corresponding non cancerous liver tissue"  
/dev\_stage="Adult"  
/lab\_host="SOLR"  
/note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 253 a 104 c 80 g 174 t  
ORIGIN

Query Match 84.0%; Score 16.8; DB 31; Length 611;  
Best Local Similarity 90.0%; Pred. No. 2.3e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 ccagatggtataagtagaac 20  
||| ||||| ||||| |||||  
Db 445 CCAATGATACATAGAGAAC 464

RESULT 3  
LOCUS A0980511 630 bp DNA GSS 29-JAN-2000  
DEFINITION RPCI-23-350G7.TJ RPCI-23 Mus musculus genomic clone RPCI-23-350G7,  
DNA sequence.  
ACCESSION A0980511  
VERSION A0980511.1 GI:6812812  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 630)  
AUTHORS Zhao S., Nieman, W., Feldblum, T., Malek, J., Shatsman, S., Akinret  
, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.  
and Fraser, C.M.  
COMMENT Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(peter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.html)  
or from Resea ch Genetics (info@resgen.com). BAC end page:  
http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 350 row: G column: 7  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source  
Location/Qualifiers

1. 630  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-350G7"  
/clone\_1lb="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site\_1:  
EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBAC3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 210 a 148 c 125 g 147 t  
ORIGIN

Query Match 84.0%; Score 16.8; DB 214; Length 630;  
Best Local Similarity 90.0%; Pred. No. 2.3e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 ccagatggtataagtagaac 20  
||| ||||| ||||| |||||  
Db 148 CCACATGATATAGAGAAC 167

RESULT 4  
LOCUS A1132934/c 710 bp mRNA EST 11-NOV-1999  
DEFINITION HA1517 Human fetal liver cDNA library Homo sapiens cDNA, mRNA  
sequence.  
ACCESSION A1132934  
VERSION A1132934.1 GI:6360250  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 710)  
AUTHORS Yu, Y., Zhang, C., Luo, L., Ouyang, S., Li, W., Wu, J., Zhou, S., Liu, M.  
and He, F.  
COMMENT Expression profile analysis of a human fetal liver cDNA library  
Unpublished (1998)  
Contact: Yongtao Yu  
Department of Hematology  
Beijing Institute of Radiation Medicine  
27 Taiping Road, Beijing 100850, P.R.China  
Tel: 0086-10-68159479  
Fax: 0086-10-68214653  
Email: yty48@yahoo.com

FEATURES  
source  
Location/Qualifiers

1. 710  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1lb="human fetal liver cDNA library"  
/tissue\_type="liver"  
/dev\_stage="fetal"  
/lab\_host="MC1061/P3"  
/note="Vector: pCDNA1"

BASE COUNT 210 a 100 c 118 g 280 t 2 others  
ORIGIN

Query Match 84.0%; Score 16.8; DB 16; Length 710;  
Best Local Similarity 90.0%; Pred. No. 2.3e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 ccagatggtataagtagaac 20  
||| ||||| ||||| |||||  
Db 140 CCAATGATACATAGAGAAC 121

RESULT 5  
LOCUS BF540037 846 bp mRNA EST 11-DEC-2000  
DEFINITION 602050669F1 NCL\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:418937 5',  
mRNA sequence.  
ACCESSION BF540037  
VERSION BF540037.1 GI:11627418  
KEYWORDS EST.  
SOURCE house mouse.

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.  
 REFERENCE 1 (bases 1 to 846)  
 AUTHORS NIH-MGC http://mhc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM9516 row: 1 column: 18  
 High quality sequence stop: 107.  
 Location/Qualifiers  
 1..846  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4189937"  
 /clone\_1lb="NCI CGAP\_SG2"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: salivary gland; Vector: pCW-SPORT6; Site\_1:  
 Not; Site\_2: Salt; cloned unidirectionally. Primer: Oligo  
 dt. Average insert size 1.3 kb. Constructed by Life  
 Technologies. Note: this is a NCI CGAP Library."  
 302 a 100 c 191 g 253 t

BASE COUNT 302 a 100 c 191 g 253 t  
 ORIGIN

Query Match 84.0%; Score 16.8; DB 149; Length 846;  
 Best Local Similarity 90.0%; Pred. No. 2.4e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ccagatgataagtagaac 20  
 1 |||||  
 Db 735 CGAGATTGTAATGAGAAC 754

RESULT 6  
 BF692672/c 894 bp mRNA EST 22-DEC-2000  
 LOCUS 602249045f1 NIH-MGC-62 Homo sapiens cDNA clone IMAGE:434291 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BF692672  
 VERSION BF692672.1 GI:11978080  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 894)  
 AUTHORS NIH-MGC http://mhc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: ATCC/DC/DTF  
 cDNA Library Preparation: CLOUTERCH Laboratories, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM1203 row: 1 column: 12  
 High quality sequence stop: 606.  
 Location/Qualifiers

FEATURES  
 source

source 1..894  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:434291"  
 /clone\_1lb="NIH-MGC-62"  
 /tissue\_type="melanotic melanoma, high MDR"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: skin; Vector: pDNR-LIB (Clontech); Site\_1:  
 SfiI (ggcgccgcgcgc); Site\_2: SfiI (ggcgccatggcc);  
 double-stranded cDNA was prepared from cell line RNA. 5'  
 and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-ATCTAGAGCGCGGCGCGACATG-3' and 3'  
 adaptor sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor  
 (where B = A, C, G and N = A, C, G, or T). Average  
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA)."  
 214 a 215 c 213 g 252 t

BASE COUNT 214 a 215 c 213 g 252 t  
 ORIGIN

Query Match 84.0%; Score 16.8; DB 151; Length 894;  
 Best Local Similarity 90.0%; Pred. No. 2.4e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ccagatgataagtagaac 20  
 1 |||||  
 Db 648 CCAGATGTAATGAGAAC 629

RESULT 7  
 BE807035/c 188 bp mRNA EST 21-NOV-2000  
 LOCUS BE807035  
 DEFINITION Gm-c1047-2584 5', mRNA sequence.  
 ACCESSION BE807035  
 VERSION BE807035.1 GI:10238147  
 KEYWORDS EST.  
 SOURCE soybean.  
 ORGANISM Glycine max  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
 Fabales; Fabaceae; Papilionoideae; Glycine.  
 REFERENCE 1 (bases 1 to 188)  
 AUTHORS Shoemaker R., Kelm P., Vodkin L., Eipelting J., Corvett V., Khanna  
 A., Bolla B., Marra M., Hillier L., Kucaba T., Martin J., Beck C.,  
 Wylie T., Underwood K., Steptoe M., Theising B., Allen M., Bowers  
 Y., Person B., Swaller T., Gibbons M., Pape D., Harvey N., Schurk  
 R., Riltter E., Kohn S., Shin T., Jackson Y., Cardenas M., McCann  
 R., Waterston R. and Wilson R.  
 Public Soybean EST Project  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Shoemaker R./Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available through: Genome Systems, Inc. 4633 World  
 Parkway Circle St. Louis, Missouri 63134 For further information  
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
 427-3324 or contact: clones@genomesystems.com or  
 info@genomesystems.com web site: www.genomesystems.com  
 Insert Length: 659 Std Error: 0.00  
 High quality sequence stop: 174.  
 Location/Qualifiers  
 1..188  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1047-2584"  
 /clone\_1lb="Gm-c1047"

FEATURES  
 source



adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 61 a 58 c 57 g 46 t

Query Match 82.0% Score 16.4; DB 11: Length 222;  
Best Local Similarity 94.4%; Pred. No. 3.5e+02;  
Matches 17: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccagatgataagtaga 18  
||| ||||| ||||| |||||  
Db 156 CCACATGCTATAGTAGA 173

RESULT 10

AA458264 242 bp mRNA EST 06-JUN-1997  
LOCUS v647c09.r1 Soares\_mammary\_gland\_NBMKG Mus musculus cDNA clone  
DEFINITION IMAGE:664496 5', mRNA sequence.

ACCESSION AA458264  
VERSION AA458264.1 GI:2180984

KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 242)  
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisels, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theisling, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HHMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (Info@image.llnl.gov) for further information.  
MGI:508584  
Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 233.

FEATURES  
SOURCE

1. 242  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:664496"  
/clone\_1lb="Soares\_mammary\_gland\_NBMKG"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAGTGGAGCGCCGCAATGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 72 a 56 c 55 g 59 t

ORIGIN

Query Match 82.0% Score 16.4; DB 7: Length 242;  
Best Local Similarity 94.4%; Pred. No. 3.5e+02;  
Matches 17: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccagatgataagtaga 18  
||| ||||| ||||| |||||  
Db 44 CCACATGCTATAGTAGA 61

RESULT 11

BF542933 250 bp mRNA EST 11-DEC-2000  
LOCUS UI-R-Y0-abm-g-12-0-UI.r1 UI-R-Y0 Rattus norvegicus cDNA clone  
DEFINITION UI-R-Y0-abm-g-12-0-UI 5', mRNA sequence.

ACCESSION BF542933  
VERSION BF542933.1 GI:11634088

KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. 1 (bases 1 to 250)  
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
COMMENT Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: mscares@blue.weeg.uiowa.edu

cDNA library preparation: M.B. Soares lab clone distribution: clones will be available through Research Genetics (www.resgen.com) This clone is also available through the I.N.A.G.E. Consortium at LNL (Info@image.llnl.gov). IMAGE ID= 1791049  
Seq primer: M13 Forward.

FEATURES  
SOURCE

1. 250  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-Y0-abm-g-12-0-UI"  
/clone\_1lb="UI-R-Y0"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-Y0 library is a subtracted library derived from an individually-tagged normalized whole-eye (minus the lens) library. The driver for the subtraction consisted of a pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, and UI-R-G1). The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-Y0) was constructed as follows: PCR amplified cDNA inserts from previous library clones from which 3' ESTs had been derived were used as a driver in a hybridization with the normalized whole-eye library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-Y0 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,

```

BASE COUNT      83 a 56 c 60 g 51 t
ORIGIN
Query Match      82.0%; Score 16.4; DB 149; Length 250;
Best Local Similarity 94.4%; Pred. No. 3.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 agatcgataagtagaac 20
        |||||||
Db      168 AGATGTTATAATAGAAC 185

RESULT 12
LOCUS      BF812297      271 bp      mRNA      EST      12-JAN-2001
DEFINITION RC2-C10090-211100-017-b01 C10090 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BF812297
VERSION      BF812297.1 GI:12141908
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 271)
AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
            Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
            Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
2020263
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC2&tl2=RC2-C10090-
            211100-017-b01&f3=2000-11-21&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 26
            Location/Qualifiers
                source          1..271
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="C10090"
                /dev_stage="Adult"
                /note="Organ: colon; ins: Vector: puc18; Site_1: SmaI;
                Site_2: SmaI; A mini-library was made by cloning products
                derived from ORESTES PCR (U.S. Letters Patent application
                No. 196,716 - Ludwig Institute for Cancer Research)
                profiles into the pUC 18 vector. Reverse transcription of
                tissue mRNA and cDNA amplification were performed under
                low stringency conditions."
BASE COUNT      72 a 60 c 49 g 90 t
ORIGIN
Query Match      82.0%; Score 16.4; DB 169; Length 271;
Best Local Similarity 94.4%; Pred. No. 3.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 agatcgataagtagaac 20

```

```

Db      128 AGATGTTATAATAGAAC 111
        |||||||
RESULT 13
LOCUS      F06213      281 bp      mRNA      EST      19-FEB-1995
DEFINITION HSC10G011 normalized infant brain cDNA Homo sapiens cDNA clone
            c-10q11, mRNA sequence.
ACCESSION  F06213
VERSION      F06213.1 GI:6770029
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 281)
AUTHORS    Aulfray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
            ,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
            Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabakchis,C. and Tessier,A.
            IMAGE: molecular integration of the analysis of the human genome
            and its expression
            C.R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
95277534
COMMENT     Contact: Genethon
            Genexpress-Genethon
            Genethon Centre de recherche sur le Genome Humain
            1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
            Tel: 33169472800
            Fax: 33160778698
            Email: genexpress@genethon.fr
            Single read.
            Genexpress library_idt: C; Genexpress_sequence_idt: ylc-10q11
            Seq primer: (-21)M13 universal.
            Location/Qualifiers
                source          1..281
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="c-10q11"
                /clone_lib="normalized infant brain cDNA"
                /sex="Female"
                /tissue_type="total brain"
                /dev_stage="3 months old"
                /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
                Site_2: NotI; sex=Female; dev_stage=3 months old;
                isolate=muscular atrophy patient; tissue_type=total brain
                ; total mRNA was oligo-(dT) primed and directionally
                cloned 5' -> 3' into the HindIII -> NotI sites of the
                lafmid BA vector. Clone library from B.Souares, Psychiatry
                Dept. Columbia University, USA. Normalization_method:
                Bento Soares, P.N.A.S in press"
BASE COUNT      92 a 52 c 65 g 70 t 2 others
ORIGIN
Query Match      82.0%; Score 16.4; DB 152; Length 281;
Best Local Similarity 94.4%; Pred. No. 3.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 agatcgataagtagaac 20
        |||||||
Db      175 AGATGTTATAATAGAAC 192

RESULT 14
LOCUS      F06677      291 bp      mRNA      EST      20-FEB-1995
DEFINITION HSC1G0091 normalized infant brain cDNA Homo sapiens cDNA clone
            c-19q09, mRNA sequence.
ACCESSION  F06677
VERSION      F06677.1 GI:672280
KEYWORDS    EST.

```



SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 291)  
AUTHORS Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes  
, M.D., Duprat, S., Houligatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F.,  
Mitchell, H., Mariage-Samson, R., Pletu, G., Pouliot, Y.,  
Sebastien-Kabackchis, C. and Tessier, A.  
IMAGE: molecular integration of the analysis of the human genome  
and its expression  
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
95277534  
JOURNAL MEDLINE  
COMMENT Contact: Genethon  
Genexpress-Genethon  
Genethon Centre de recherche sur le Genome Humain  
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE  
Tel: 33169472800  
Fax: 33160778698  
Email: genexpress@genethon.fr  
Single read.  
Genexpress\_library\_id: C: Genexpress\_sequence\_id: y1c-19909  
Seq primer: (-21)M13-universal.  
Location/Qualifiers  
1..291  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="normalized infant brain CDNA"  
/sex="Female"  
/tissue\_type="total brain"  
/dev\_stage="3 months old"  
/note="Organ: brain; Vector: lafmid BA; Site:1: HindIII;  
Site:2: NotI; sex:Female; dev\_stage=3 months old;  
Isolate=muscular atrophy patient; tissue\_type=total brain  
; total mRNA was oligo-(dT) primed and directionally  
cloned 5' -> 3' into the HindIII -> NotI sites of the  
lafmid BA vector. Clone library from B.Saeres, Psychiatry  
Dept. Columbia University, USA. Normalization\_method:  
Bento Soares, P.N.A.S in press"

BASE COUNT 94 a 53 c 68 g 72 t 4 others  
ORIGIN

Query Match 82.0%; Score 16.4; DB 152; Length 291;  
Best Local Similarity 94.4%; Pred. No. 3.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 agatgataagtagaac 20  
|||||  
DB 175 AGATGCTATAATACGAC 192

RESULT 15  
AW313588 298 bp mRNA EST 08-JAN-2001  
LOCUS 8892 MARC 2BOV Bos taurus CDNA 5', mRNA sequence.  
ACCESSION AW313588  
VERSION AW313588.1 GI:6742808  
KEYWORDS EST.  
SOURCE COW.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 298)  
AUTHORS Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M.,  
Bennett, G.A., Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Laegreid  
, W.W. and Keele, J.W.  
Design and use of four pooled tissue normalized cDNA libraries for  
EST discovery in cattle  
Unpublished (2000)  
JOURNAL COMMENT Contact: Smith TP

USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 20  
and -mismatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACAT  
BACKWARD: GTTTCACGTCACGACG  
Plate: 142 row: K column: 5  
Seq primer: ATTAGTGACACTATAG.  
Location/Qualifiers  
1..298  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 2BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site:1: XbaI; Site:2: XhoI;  
library made from pooled tissue from testis, thymus,  
semitepidonous muscle, longissimus muscle, pancreas,  
adrenal, and endometrium."

BASE COUNT 98 a 55 c 72 g 73 t  
ORIGIN

Query Match 82.0%; Score 16.4; DB 113; Length 298;  
Best Local Similarity 94.4%; Pred. No. 3.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 agatgataagtagaac 20  
|||||  
DB 175 AGATGCTATAATACGAC 192

Search completed: June 8, 2001, 05:10:38  
Job time: 4328 sec

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2001, 06:02:45 ; Search time 1501.23 Seconds  
(without alignments) 196.477 Million cell updates/sec

Title: US-09-507-242-6

Perfect score: 20  
Sequence: 1 caatgacgggaaacgttcg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1283235 segs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenDbml:\*  
1: gb\_da1:\*  
2: gb\_da2:\*  
3: gb\_da3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_om:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: em\_ba1:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_inv:\*  
22: em\_htg\_hum1:\*  
23: em\_htg\_hum2:\*  
24: em\_htg\_hum3:\*  
25: em\_htg\_hum4:\*  
26: em\_htg\_hum5:\*  
27: em\_htg\_hum6:\*  
28: em\_htg\_hum7:\*  
29: em\_htg\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_other:\*  
33: em\_htg\_rtd:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_om:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_v1:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_v11:\*  
59: gb\_v12:\*  
60: gb\_v13:\*  
61: gb\_v14:\*  
62: gb\_v15:\*  
63: gb\_v16:\*  
64: gb\_v17:\*  
65: gb\_v18:\*  
66: gb\_v19:\*  
67: gb\_v20:\*  
68: gb\_v21:\*  
69: gb\_v22:\*  
70: gb\_v23:\*  
71: gb\_v24:\*  
72: gb\_v25:\*  
73: gb\_v26:\*  
74: gb\_v27:\*  
75: gb\_v28:\*  
76: gb\_v29:\*  
77: gb\_v30:\*  
78: gb\_v31:\*  
79: gb\_v32:\*  
80: gb\_v33:\*  
81: gb\_v34:\*  
82: gb\_v35:\*  
83: gb\_v36:\*  
84: gb\_v37:\*  
85: gb\_v38:\*  
86: gb\_v39:\*  
87: gb\_v40:\*  
88: gb\_v41:\*  
89: gb\_v42:\*  
90: gb\_v43:\*  
91: gb\_v44:\*  
92: gb\_v45:\*  
93: gb\_v46:\*  
94: gb\_v47:\*  
95: gb\_v48:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.8	84.0	980	9 AX065753	AX065753 Sequence
2	16.8	84.0	1404	9 AX065751	AX065751 Sequence
3	16.8	84.0	4161	2 BACGIDA	M65289 Bacillus st
4	16.8	80.0	674	1 AF083905	AF083905 Salmonell
5	16.8	80.0	12099	1 AF083912	AF083912 Salmonell
6	16.8	80.0	44237	6 CELY6584BR	AC024847 Methanoba
7	16.8	80.0	267118	60 AC006889	AC006889 Caenorhab
8	15.8	79.0	14625	1 AE005077	AE005077 Haibacte
9	15.8	79.0	22440	3 MTV006	AE021006 Mycobacte
10	15.8	79.0	22480	5 CELF44E7	AF016421 Caenorhab
11	15.8	79.0			

12	15.8	79.0	31205	5	CELC06G1	U41014 Caenorhabdi
C 13	15.8	79.0	47436	78	AC087901	AC087901 Mus muscu
C 14	15.8	79.0	72593	67	AC021519	AC021519 Homo sapi
C 15	15.8	79.0	110000	84	LMFLCHR34_1	Continuation (2 of
C 16	15.8	79.0	174801	64	AC014376	AC014376 Drosophi
C 17	15.8	79.0	174823	61	AC009908	AC009908 Homo sapi
C 18	15.8	79.0	176340	78	AF216671	AF216671 Homo sapi
C 19	15.8	79.0	179989	64	AC016451	AC016451 Homo sapi
C 20	15.8	79.0	285978	4	AE003551	AE003551 Drosophi
C 21	15.8	79.0	303650	2	AP001519	AP001519 Bacillus
C 22	15.4	77.0	61920	66	AC021270	AC021270 Homo sapi
C 23	15.4	77.0	82315	12	AB006707	AB006707 Arabidops
C 24	15.4	77.0	100727	63	AC013831	AC013831 Drosophi
C 25	15.4	77.0	104204	13	ATT9G5	AL132964 Arabidops
C 26	15.4	77.0	107931	12	AC007478	AC007478 Arabidops
C 27	15.4	77.0	133195	12	AC007123	AC007123 Arabidops
C 28	15.4	77.0	134610	76	AC079302	AC079302 Homo sapi
C 29	15.4	77.0	194041	76	AC079349	AC079349 Homo sapi
C 30	15.4	77.0	349061	3	NMA22491	AL167353 Neisseria
C 31	15.2	76.0	126	54	G54042	G54042 1562 Human
C 32	15.2	76.0	490	3	UBA244349	AJ244349 unculture
C 33	15.2	76.0	490	3	STU09529	U09529 Salmonella
C 34	15.2	76.0	1073	2	AS0252625	AJ252625 Agriicultu
C 35	15.2	76.0	1073	2	AE020540	AE020540 Mus muscu
C 36	15.2	76.0	1103	94	AK025868	AK025868 Homo sapi
C 37	15.2	76.0	1705	89	MAV250022	AJ250022 Mycobacte
C 38	15.2	76.0	1856	92	HSM802238	AL137722 Homo sapi
C 39	15.2	76.0	2856	92	ECOXYLABA	K01996 E.coli xyla
C 40	15.2	76.0	4176	2	ECOXYLABA	X04691 E.coli gen
C 41	15.2	76.0	4176	2	ECOXYLABA	M17381 Streptococ
C 42	15.2	76.0	4995	3	STRCTFI	E03388 Xylose isom
C 43	15.2	76.0	7801	10	E03388	AF098506 Caenorhab
C 44	15.2	76.0	8770	5	CELH29C22	Z68886 Human DNA s
C 45	15.2	76.0	10589	92	HS121F12C	

## ALIGNMENTS

RESULT 1						
AX065753	980 bp	DNA				
DEFINITION	Sequence 97 from Patent WO0100842.	PAT	24-JAN-2001			
ACCESSION	AX065753					
VERSION	AX065753.1					
KEYWORDS	GI:12543465					
SOURCE						
ORGANISM	Corynebacterium glutamicum.					
	Corynebacterium glutamicum.					
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;					
	Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;					
	Corynebacterium.					
	1 (bases 1 to 980)					
REFERENCE	Pompeius, M., Kroeger, B., Schroeder, H., Zelder, O. and Habermayer, G.					
AUTHORS	Corynebacterium glutamicum genes encoding proteins involved in					
TITLE	homeostasis and adaptation					
JOURNAL	Patent: WO 0100842-A 97 04-JAN-2001;					
	Basf AKTIENGESellschaft (DE)					
FEATURES	Location/Qualifiers					
source	1..980					
	/organism="Corynebacterium glutamicum"					
	/db_xref="taxon:1718"					
	<1..960					
	/note="FRXA00499"					
	/codon_start=1					
	/transl_table=11					
	/protein_id="CAC26116.1"					
	/db_xref="GI:12543465"					
	/translation="GNDILEKPTAEVVEYLISHLRADGIVRDAEALRKHLGVNOMNLG					
	OSFGFTTLYLHLSRHADSLDNVEITGSLAIDRAPADVANCYNNRRRSSEERYREF					
	OLREFFGTVNLRAGEIYLPTEGVSETRLSIGHLGSDNGMFLYLNLELDPNSN					
	AVYDLAGLLPGKRNPIYVLIHSSYADGVYTNMAAEVLPEDFEDPTLLTGEIVF					
	QEWTDVPSLTKPRKDALALAOQEMFKLYDAKALENSQAKGAAYVANDVFVPVDYSL					
	ETRAQHLPGVOLFTTSQEHNGLRASSGAVLKLHFLDLAHGREVR"					

## CDS

/note="FRXA00499"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAC26116.1"  
 /db\_xref="GI:12543465"  
 /translation="GNDILEKPTAEVVEYLISHLRADGIVRDAEALRKHLGVNOMNLG  
 OSFGFTTLYLHLSRHADSLDNVEITGSLAIDRAPADVANCYNNRRRSSEERYREF  
 OLREFFGTVNLRAGEIYLPTEGVSETRLSIGHLGSDNGMFLYLNLELDPNSN  
 AVYDLAGLLPGKRNPIYVLIHSSYADGVYTNMAAEVLPEDFEDPTLLTGEIVF  
 QEWTDVPSLTKPRKDALALAOQEMFKLYDAKALENSQAKGAAYVANDVFVPVDYSL  
 ETRAQHLPGVOLFTTSQEHNGLRASSGAVLKLHFLDLAHGREVR"

BASE COUNT	215 a	267 c	264 g	234 t
ORIGIN				
Query Match	84.0%; Score 16.8; DB 9; Length 980;			
Best Local Similarity	90.0%; Pred. No. 2.8e+02;			
Matches 18; Conservative	0; Mismatches 2; Indels 0; Gaps 0;			

QY 1 caatgacggagaacgttcg 20  
 ||||| ||||| ||||| |||||  
 Db 307 CAATTCGGGAACTTCCG 326

RESULT 2						
AX065751	1404 bp	DNA				
DEFINITION	Sequence 95 from Patent WO0100842.	PAT	24-JAN-2001			
ACCESSION	AX065751					
VERSION	AX065751.1					
KEYWORDS	GI:12543463					
SOURCE						
ORGANISM	Corynebacterium glutamicum.					
	Corynebacterium glutamicum.					
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;					
	Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;					
	Corynebacterium.					
	1 (bases 1 to 1404)					
REFERENCE	Pompeius, M., Kroeger, B., Schroeder, H., Zelder, O. and Habermayer, G.					
AUTHORS	Corynebacterium glutamicum genes encoding proteins involved in					
TITLE	homeostasis and adaptation					
JOURNAL	Patent: WO 0100842-A 95 04-JAN-2001;					
	Basf AKTIENGESellschaft (DE)					
FEATURES	Location/Qualifiers					
source	1..1404					
	/organism="Corynebacterium glutamicum"					
	/db_xref="taxon:1718"					
	101..1384					
	/note="RXN00499"					
	/codon_start=1					
	/transl_table=11					
	/protein_id="CAC26115.1"					
	/db_xref="GI:12543464"					
	/translation="MVGVSFTRPARNIGSWTKTLGSLQLEITLTLPEDVADEETID					
	VFARIATRVGGEDLPYLVFLQGGPGNEARFSLNPINPWLVALEEVVAVLDORGT					
	GRSTVAGNDILEKPTAEVVEYLISHLRADIVDAEALRKHLGVNOMNLGOSFGFTT					
	LHLSRHADSLDNVEITGSLAIDRAPADVANCYNNRRRSSEERYREFRPPQRETFRG					
	LVNBARAGEIYLPTEGVSETRLSIGHLGSDNGMFLYLNLELDPNSNAVVDLAG					
	ILPGKRNPIYVLIHSSYADGVYTNMAAEVLPEDFEDPTLLTGEIVFQWMDPTVP					
	SLKPRKDALALAOQEMFKLYDAKALENSQAKGAAYVANDVFVPVDYSLTRAQHLPG					
	VOLFITSQEHNGLRASSGAVLKLHFLDLAHGREVR"					

## CDS

BASE COUNT	302 a	384 c	384 g	334 t
ORIGIN				
Query Match	84.0%; Score 16.8; DB 9; Length 1404;			
Best Local Similarity	90.0%; Pred. No. 2.8e+02;			
Matches 18; Conservative	0; Mismatches 2; Indels 0; Gaps 0;			

QY 1 caatgacggagaacgttcg 20  
 ||||| ||||| ||||| |||||  
 Db 731 CAATTCGGGAACTTCCG 750

RESULT 3						
BACGIDA/c	BACGIDA	4161 bp	DNA	BCT	26-APR-1993	
DEFINITION	Bacillus stearothermophilus	glycerol dehydrogenase (proposed gld)				
ACCESSION	gene, complete cds.					
VERSION	M65289					
KEYWORDS	glycerol dehydrogenase.					
SOURCE	Bacillus stearothermophilus (sub-species nondiastaticus) (library: lambda, EcoRI fragments) DNA.					
ORGANISM	Bacillus stearothermophilus					

Bacteria; Firmicutes; Bacillus/Clostridium group;  
Bacillus/Staphylococcus group; Bacillus.  
1 (bases 1 to 4161)  
Mallinder, P. R., Pritchard, A. and Molt, A.  
Cloning and characterization of a gene from Bacillus  
stearothermophilus var. non-diastaticus encoding a glycerol  
dehydrogenase  
Gene 110, 9-16 (1992)  
92184120

JOURNAL  
MEDLINE  
FEATURES  
SOURCE

Location/Qualifiers  
1. .4161  
/organism="Bacillus stearothermophilus"  
/sub-species="nondiastaticus"  
/db\_xref="taxon:1422"  
/tissue\_1lb="lambda, EcoRI fragments"  
complement<1. .292)  
/note="ORF1"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AA022476.1"  
/db\_xref="GI:551708"  
/translation="MDILKTEKHREMERKMGTEFAVELEILKEKPWIAOSASRV  
YNMKDGIEIEINRKRKFFNQEIFGLDELRLVEEYFPAAKRLDYRKRI"  
742. .1854  
/gene="gld"  
742. .1854  
/gene="gld"  
/EC\_number="1.1.1.6"  
/codon\_start=1  
/transl\_table=1  
/product="glycerol dehydrogenase"  
/protein\_id="AA022477.1"  
/db\_xref="GI:142978"  
/translation="MAAEVFTSPAKYVOGKNVITKIANLEGKNTVIADEIVMK  
IAGTIVNELKKNIAAEVYFSEASNEVERIANIRKKAALVIVGSGKITDPA  
KAYADEDAIVITPAISTADPTSAISVITSDGVEFSTFVKNPDLVLDKITA  
NAPRRLASGIADALATWVEARSVIKSGGIMAGCIPPIAAEAIAKEQTLFKYGL  
AYESKAKVVPALAEVAVENTLISGLFESGGAAAHAINGFALGELIHLTHGE  
KAVGTLVQLALEHSOQEIERYTELVSLLPVLIEDIKLKDASREDILKAKATA  
EGETIHNAFNTADVADAIAPADQYAAKAYEKHKR"  
complement(2044. .2517)  
/note="ORF3 is homologous to an ORF downstream of the spot  
gene of E. coli; ORF3"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AA022478.1"  
/db\_xref="GI:142979"  
/translation="MALHVLVPEPEIPANTGNIAKTCATNTALHLIRPLGFTDDKA  
LRAGLDYWEFVNIIHYDSLDFEQONONGDFYITKGRKRYTSYDSDSKELFFI  
FGRETTGLPKELAEINEDRCLRIPMTENVALNLSNTAAIIVYEAALROOSFHGUS"  
complement(2550. .3455)  
/note="ORF4"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AA022479.1"  
/db\_xref="GI:142980"  
/translation="MDVOAYVYKINENKLOFLVKNHVNHRROOSSELEWCEKKKELAAK  
RKAEIVATCLIKPKQIODEATVVDYICDRLIYKQDPLFIEIEIGERAKKSNK  
HLISDEVKIKDCFKDEEDDSNEEMHKKDERVSFOYDLKLSVOYAEKMMASRNPNP  
YFDVNCINIVSQCLRAGDAPMRGYPNRSNGWMLNNMNSYSVAHALMLLYLSKTSK  
LRAEVSSPELKLPGVICYDPOGDRDHSFTVAAKDSMPLVNNAOTSRRMAYWS  
VEDSTAYTENIOYKFFHIIIDNEDS"  
complement(3511. .4161)  
/partial  
/note="ORF5; This ORF includes a region (aa23-103)  
containing a potential iron-sulphur centre homologous to a  
region of Rhodospirillum rubrum and Chromatium vinosum;  
putative"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AA022480.1"  
/db\_xref="GI:142981"  
/translation="EFGSYVLGEMITNIPPPSPIEDRCGSCNICVDSCTGALVO

GGOLDSKCIAFLAQTKGLDEPRKIGNRLYGCOTQLVCPENKKGDFHIHSEMER  
DPEIAKPKLPLHLISNREFKEFGHISGRKCKPIORNAIILAHYKDTALPDLI  
KLHEDSPRIIRGTAAWALIGRIGDOSVLEFELEKALEKRQDEVKNEIVKGIEMLKA"  
BASE COUNT 1235 a 849 c 790 g 1287 t  
ORIGIN

Query Match 84.0%; Score 16.8; DB 2; Length 4161;  
Best Local Similarity 90.0%; Pred. No. 2.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 caatgacgggaaacgtccg 20  
Db 2141 CAATGACGGAAACGTTCCG 2122

RESULT 4  
AF083905/c 674 bp DNA BCT 08-APR-1999  
LOCUS  
DEFINITION  
Salmomella enterica VI strain RK3057 major pilin protein Fima  
(fima) and fimbriin-like protein Fimi (fimi) genes, partial cds.  
AF083905  
ACCESSION  
AF083905.1 GI:4574180  
KEYWORDS  
SOURCE  
ORGANISM  
Salmomella enterica VI.  
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Salmomella.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
SOURCE  
1. .674  
/organism="Salmomella enterica VI"  
/strain="RK3057"  
/sub-species="VI"  
/db\_xref="taxon:59207"  
<1. .486  
/gene="fima"  
/note="major pilin gene"  
<1. .486  
/gene="fima"  
/codon\_start=1  
/transl\_table=1  
/product="major pilin protein Fima"  
/protein\_id="AAD23933.1"  
/db\_xref="GI:4574181"  
/translation="PTPVSVSGGTHFEKGLVNAACAVSTKSADQVTTLGQRTASFT  
AIGNTTAQVPTIVLDCDPPKAAATAVAAFSGQDNFTNNLLAASANSATATGVGI  
EILDNTSSPLKPDGATFSAKOSIVEGTTLFTARYKATAAATTGGANADATFIMKY  
E"  
561. .>674  
/gene="fimi"  
561. .>674  
/gene="fimi"  
/codon\_start=1  
/transl\_table=1  
/product="fimbriin-like protein Fimi"  
/protein\_id="AAD23934.1"  
/db\_xref="GI:4574182"  
/translation="MIRKGAIVGLITFPVIAQSVIVSGRVHLRGLVNG"

BASE COUNT 161 a 182 c 190 g 141 t  
ORIGIN

Query Match 80.0%; Score 16; DB 1; Length 674;  
Best Local Similarity 100.0%; Pred. No. 7.7e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caatgacgggaacct 16  
|||||  
Db 615 CAATGACGGGAACCT 600

RESULT 5  
AF083912/c  
LOCUS AF083912 674 bp DNA BCT 08-APR-1999  
DEFINITION *Salmonella enterica* VI strain RKS2995 major pilin protein FimA (fima) and fimbria-like protein FimI (fimI) genes, partial cds.  
ACCESSION AF083912  
VERSION AF083912.1 GI:4574201  
KEYWORDS  
SOURCE *Salmonella enterica* VI.  
ORGANISM *Salmonella enterica* VI.  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; *Salmonella*.

REFERENCE 1 (bases 1 to 674)  
AUTHORS Boyd, E.F. and Hartl, D.L.  
TITLE Analysis of the type 1 pilin gene cluster fim in *Salmonella*: its distinct evolutionary histories in the 5' and 3' regions  
JOURNAL J. Bacteriol. 181 (4), 1301-1308 (1999)  
MEDLINE 99138752  
REFERENCE 2 (bases 1 to 674)  
AUTHORS Boyd, E.F.  
TITLE Direct Submission  
JOURNAL Submitted (13-AUG-1998) Organismic & Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA  
FEATURES  
source Location/Qualifiers  
1..674  
/organism="Salmonella enterica VI"  
/strain="RKS2995"  
/sub-species="VI"  
/db\_xref="taxon:59207"  
<1..486  
/gene="fima"  
/note="major pilin gene"  
<1..486  
/gene="fima"  
/codon\_start=1  
/transl\_table=1  
/product="major pilin protein FimA"  
/protein\_id="AAD23947.1"  
/db\_xref="GI:4574202"  
/translation="PTPEVSYSGGTHIEGKLVNACAVSTKSADQVTLGQYRTAST  
AIGTTTAQVPTIVLNDCKPRKATATAVAATVSGADNTNNLLAVSSADNSATATGAGI  
EILDNTSSPLKPGATFSAKQSLVEGTNLTFRATARKATATAATTPGOANADATFIMKY  
E"  
561..>674  
/gene="fimI"  
561..>674  
/gene="fimI"  
/codon\_start=1  
/transl\_table=1  
/product="fimbria-like protein FimI"  
/protein\_id="AAD23948.1"  
/db\_xref="GI:4574203"  
/translation="MIRKSALVGLITFPVIAOSVIVESGRVHLGSKLVNG"  
BASE COUNT 162 a 180 c 190 g 142 t  
ORIGIN

Query Match 80.0%; Score 16; DB 1; Length 674;  
Best Local Similarity 100.0%; Pred. No. 7.7e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caatgacgggaacct 16  
|||||  
Db 615 CAATGACGGGAACCT 600

RESULT 6  
AE000941/c  
LOCUS AE000941 12099 bp DNA BCT 15-NOV-1997  
DEFINITION *Methanobacterium thermoautotrophicum* from bases 1724442 to 1736540 (section 147 of 148) of the complete genome.  
ACCESSION AE000941  
VERSION AE000941 AE000066  
KEYWORDS  
SOURCE AE000941.1 GI:2623025  
ORGANISM *Methanobacterium thermoautotrophicum*.  
Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;

REFERENCE 1 (bases 1 to 12099)  
AUTHORS Smith, D.R., Doucette-Stamm, L.A., Delouhery, C., Lee, H., Dubois, J., Aldridge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothler, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Reeve, J.N. et al.  
TITLE Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics  
J. Bacteriol. 179 (22), 7135-7155 (1997)  
MEDLINE 96037514  
REFERENCE 2 (bases 1 to 12099)  
AUTHORS Smith, D.R.  
TITLE Direct Submission  
JOURNAL Submitted (10-AUG-1997) Genomics and Technology Development, Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154-8448, USA  
FEATURES  
source Location/Qualifiers  
1..12099  
/organism="Methanobacterium thermoautotrophicum"  
/strain="delta H"  
/db\_xref="taxon:145262"  
/clone="MTH"  
complement(64..960)  
/gene="MTH1893"  
complement(64..960)  
/gene="MTH1893"  
/note="Function Code:12.11 - Cell Processes, Detoxification; similar to, pir.LN:A64356 AC:A64356, p()=1.8E-36, pid=198"  
/codon\_start=1  
/transl\_table=1  
/product="cation efflux system protein (zinc/cadmium)"  
/protein\_id="AAB86353.1"  
/db\_xref="GI:2623026"  
/translation="MESSEKIRILGKRAAFGIGCNVLTSLNPLVIGISSGVALLVAEA  
AHTLSVLTIVITVIGFRIGQRPDPKHPGHRARALVGLVAVVFLGIIITSLSEA  
YRFLTELAPPDYATLMAFGGITIANMTYLRIGERINSPAIYADAOHQVDEDS  
EIAIMLGAVAGSHLGFELDFLAVIIVIAVLKTAIFDVGRENVNIIIGAVSPIMEDI  
EKASMSYDVGKGLHEVRIINHEGPAVSVDLHIEVDGDMRLDARIHADVERVIEDID  
IIRIVNVHVCPEKSCLEMS"  
complement(962..2086)  
/gene="MTH1894"  
complement(962..2086)  
/gene="MTH1894"  
/note="Function Code:5.01 - L-Amino Acid Metabolism, Alanine--aspartate and glutamate metabolism; similar to, sp.LN:AAI-METUA AC:060317, p()=6.1E-82, pid=468"  
/codon\_start=1  
/transl\_table=1  
/product="aspartate aminotransferase homolog"  
/protein\_id="AAB86354.1"  
/db\_xref="GI:2623027"  
/translation="MIMKRTASRVKDTQSEIRKIEFVADPTINIGIGPDSVPDH  
VREAVKADVDEGLTHITSNKGEELEBALDKSEKRVHAEPESITVYGAEEALFM  
CTGALLDIGHALIPDPGLSYDACVRLSGAVSIPLVLSDEGSSPESVSLITQD  
TVRIINSPNPSTGSGKDVKGIAEIAEDNLIITISDEYKRIIYDGKHSPPAQT  
DNALIVNGFSKTYAMTGLRIGYVAGCEDITEELKLVQYVTAAPSISOYALAAIAG  
PNCVVKMVDPEFRRLDMRSLITDMGLECVLPGGAFYMPVYAGDSEEFKLSLEAGV  
AVVPGSAFGNEGKGIIRMSATVTSIELIEAMEKLVYCGV"

```

gene      complement(2121..3626)
           /gene="MTH1895"
CDS       complement(2121..3626)
           /gene="MTH1895"
           /note="Function Code:14.01 - Unknown, Conserved protein;
           similar to, gp:GI:91652365 LN:D90905, p(-)=9.8E-41,
           pid=35%"
           /codon_start=1
           /transl_table=11
           /product="conserved protein"
           /protein_id="AAB86358.1"
           /db_xref="GI:2623028"
           /translation="MRAVASCYSPSPYSTAMSSLGHIITHHFINSREDTWCERTIHPHR
           SFESGSPRLDFDVVSTIHYEDYFRIPAMFRCGLERDERGRLVIAGPCITSN
           PLPSDFIDLEIIGEAEPVLDMDRCLEDDPRDEIDFRIDRGVYVDPSPAEIV
           ENMEACHPVQIIVPTDDRSITPRLSGSLFSLGSGRCGRCGMSGLYLRPKETS
           LRKIDISEGCEANTGYSKSLTGAASDYSRIDELCNELMGLMVMSPSLRSEVT
           PELMALIRGLKRTITMAPESTMKLRDLNKPITDSMFETKRALEGLVAKMTFLI
           GAGESREDEEMARLMDLSEIRGRVSEFVNLPKPHPLQWEGYDAREMKQIR
           FLGRVLVGLPYRMGSPRGGLIQHVLSTGSPFISGELIERASLSGVPRVREMAHAGRDP
           CARLPMHNISVGLREDFLMEBYIKMKDDRLTPMCEESGCGCMDSCLSRH"
           3797..4312
           /gene="MTH1896"
           3797..4312
           /gene="MTH1896"
           /note="Function Code:14.01 - Unknown, Conserved protein;
           similar to, p1r:LN:E64428 AC:E64428, p(-)=3.3E-26, pid=36%"
           /codon_start=1
           /transl_table=11
           /product="conserved protein"
           /protein_id="AAB86356.1"
           /db_xref="GI:2623029"
           /translation="MFERKDDTKRRKYSLVAVGTFEDRFHKGRRLLDEAFRGVETMI
           GVTSDERFAAGSEIGEPSCVYRMKMLEYLRDKDDYHMRLLDDPYGTGTEAFPAIV
           VSRTEPVAREINIRNRGRFELDIITIDVNMDDGIPISSTRIRRCEIDPMGHIIR
           RIRGALRRRR"
           4334..4939
           /gene="MTH1897"
           4334..4939
           /gene="MTH1897"
           /note="Function Code:14.01 - Unknown, Conserved protein;
           similar to, sp:LN:Y261_METUA AC:Q57709, p(-)=1.8E-27,
           pid=33%"
           /codon_start=1
           /transl_table=11
           /product="conserved protein"
           /protein_id="AAB86357.1"
           /db_xref="GI:2623030"
           /translation="MGCTFWYDLDLHIQVKEQSDSVFYMKVYVSGNPVKVRATEN
           VLGLIGNVEYRGVEVSGVDPQVGLSEETVRGAVNRRARAFRCDELSSVGSGLHRY
           PETTGEVDLQWCAIYDGEHITTLGVSAGFEYPPMVVEVLAGREVGDVMDLTGVDEL
           GRKGAVSPLSGGLDRGTNEGCVLMAIMPRMNPISLGLK"
           5152..7170
           /gene="MTH1898"
           5152..7170
           /gene="MTH1898"
           /note="Function Code:11.04 - Cell envelope, Murein
           sacculus and peptidoglycan; similar to, p1r:LN:D64490
           AC:D64490, p(-)=2.6E-34, pid=31%"
           /codon_start=1
           /transl_table=11
           /product="oligosaccharyl transferase"
           /protein_id="AAB86358.1"
           /db_xref="GI:2623031"
           /translation="MROYMRFLLVAVIFVFGFSIRMLNAGYDDPLTYTDDGLQYFQF
           SDSITYNRTLENLAIRGIPGDVAVIGKIMDLHSTYPPGVLDYPAIGILLAVILHAIIT
           YGAVPSLRAVCLIMPLAPLAVIGICFTYPRQGDIPATIAGLILVSAQYVAVY
           YGLYDMLILITLSMAVLEIADAREVNVPSAIGGFIIFLESTWSGQWVYV
           TLVFEIILITSGSRKPLRNLTNGIPNGISVYPFNENLELIKILMPLHIDLV
           DKSTVWPDSYTYVAELPAYMDKVGHISPATVYGAADIPAYMYLRRKSGVGLP
           VIFISWIATGALMLGCIREFVMLINPLVSACIFGILDCSRKEFRNWKGLFI
           LIIFTVAQLAISGNMVAADLKPGYNDFEESAHIAENTPDAVITTEMWSGHVYASE
           ADRPVYVDGLAYIETLPARGVWYSSMDPEIPTTARDYWINLSLSENRTLSTINLR

```

```

gene      MLANSGDVAYILKNQGSSEVQAVRINNELGLNRESAMDVLMHRYGSESEALRILN
           MTHPMTSRKIIIVVYVEKMIIGLIASGSDGMDPAGYSLIRKDVIEEELINSSGNWT
           VLEETDGSYRINRKYRKSLLLEIMGVNDENSLVYRNKQIKVYIINS"
           7530..7703
           /gene="MTH1899"
           7530..7703
           /gene="MTH1899"
           /note="Function Code:14.00 - Unknown, ; similar to,
           sp:LN:YWLA_BACSU AC:P39150, p(-)=0.98, pid=15%"
           /codon_start=1
           /transl_table=11
           /product="unknown"
           /protein_id="AAB86359.1"
           /db_xref="GI:2623032"
           /translation="MYSAPCPASSKFNFLYTSINIHQFWLATLICVYPLINEADNT
           LVILCSHQYHI"
           7735..8286
           /gene="MTH1900"
           7735..8286
           /gene="MTH1900"
           /note="Function Code:14.00 - Unknown, ; similar to,
           sp:LN:IATP_PICUA AC:P09940, p(-)=0.999, pid=06%"
           /codon_start=1
           /transl_table=11
           /product="unknown"
           /protein_id="AAB86360.1"
           /db_xref="GI:2623033"
           /translation="MMPKRYNIDRYILEVLOGLDSRYELGERIRSEIFAVTDTIN
           EAFIRLKAASRTIYTGVDLYGVDYGVDEROSLKPQIGVIAQORDPVENLLIRLESE
           NHHSESALNKLRIKIPMTKGEELGVADAGISTIVNELSIDODOKRIITOKLVAALS
           DEDDAPEQLRHLIYFEIRAGNF"
           8626..10023
           /gene="MTH1901"
           8626..10023
           /gene="MTH1901"
           /note="Function Code:14.00 - Unknown, ; similar to,
           p1r:LN:S45091 AC:S45091, p(-)=0.21, pid=06%"
           /codon_start=1
           /transl_table=11
           /product="unknown"
           /protein_id="AAB86361.1"
           /db_xref="GI:2623034"
           /translation="MTYILITITAGIWMIFAGLYMSLGFATPPDIPMFFLILFVPG
           LIITLFGKVDLKSLSKASEPKDKKRIKAGAPVAEGGKQDRPGVSGIVYGE
           KTEAPKPLKSTGKYQIRKPGVYKVRTSDEIKSRAIKPTVYTESDESILGAGE
           PSAASRIKPEISELKEIIGELPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEK
           PEPEAPPSREPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEK
           DNSYTERLQRLKOEYIENVDDVEDLEDRILDSRKGAINRIATREPTSTWSVSD
           VQDAMRETIITAAEEVYVLMYPMIRNIDVSLKFKMDIESRLIIGESLDDSAVELIK
           VLMNNVQIRTMPHITHVAVAADKNGLIISTDPIYESFEVGIYKDKKSISEIKKLF
           EEAMELSNEINLEGV"
           10023..10715
           /gene="MTH1902"
           10023..10715
           /gene="MTH1902"
           /note="Function Code:14.01 - Unknown, Conserved protein;
           similar to, p1r:LN:B64445 AC:B64445, p(-)=6.3E-48, pid=46%"
           /codon_start=1

Query Match      80.0%; Score 16; DB 1; Length 12099;
Best Local Similarity 100.0%; Pred. No. 6.5e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1      caatagcggaagcgt 16
Db      10620      CAATGACGGGAACGT 10605

RESULT 7
LOCUS   CELY65B4B/c 44237 bp      DNA      28-MAR-2000
DEFINITION Caenorhabditis elegans cosmid Y65B4B/c, complete sequence.
ACCESSION AC024847

```

VERSION AC024847.1 GI:7140409  
 KEYWORDS HTG.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis. 1 (bases 1 to 44237)  
 REFERENCE 1 The C. elegans Genome Sequencing Consortium, Washington University Genome Sequencing Center, St. Louis U.S.A. and the Sanger Centre, Hinxton, U.K./C.  
 TITLE Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium Science 282 (5396), 2012-2018 (1998)  
 JOURNAL Erratum: [[published errata appear in Science 1999 Jan 1;283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep 3;285(5433):1493]]  
 REMARK 2 (bases 1 to 44237)  
 3;285(5433):1493]]  
 REFERENCE 2 (bases 1 to 44237)  
 AUTHORS Waterston,R.H.  
 JOURNAL Direct Submission  
 Submitted (01-MAR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 3 (bases 1 to 44237)  
 AUTHORS Waterston,R.  
 JOURNAL Direct Submission  
 Submitted (28-MAR-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT Submitted by:  
 Genome Sequencing Center  
 Department of Genetics, Washington University,  
 St. Louis, MO 63110, USA, and  
 Sanger Centre, Hinxton Hall  
 Cambridge CB10 1RO, England  
 e-mail: rwenematode.wustl.edu and jesusanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats: all regions were covered by sequence from more than one m13 subclone.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

FEATURES

source Location/Qualifiers

1..44237  
 /organism="Caenorhabditis elegans"  
 /strain="Bristol N2"  
 /db\_xref="taxon:6239"  
 /chromosome="I"  
 /clone="Y65B4BR"  
 misc\_feature 1386..1387  
 /note="5'UTR trans-splice site: see ESTs yk342b3.5, yk62a2.5 and yk516d1.5"  
 gene 1404..2089  
 /gene="Y65B4BR.5"  
 join(1404..1445,1495..1902,1952..2089)  
 /gene="Y65B4BR.5"  
 CDS join(1404..1445,1495..1902,1952..2089)  
 /note="contains similarity to alpha-NAC, non-muscle form proteins; coded for by C. elegans cDNA yk410q1.5; coded for by C. elegans cDNA yk361f9.5; coded for by C. elegans cDNA yk334b12.3; coded for by C. elegans cDNA yk655a1.3; coded for by C. elegans cDNA yk644h2.3; coded for by C.

elegans cDNA yk613d2.3; coded for by C. elegans cDNA yk678f10.3; coded for by C. elegans cDNA yk424f8.3; coded for by C. elegans cDNA yk584g3.3; coded for by C. elegans cDNA yk673h6.3; coded for by C. elegans cDNA yk117f10.3; coded for by C. elegans cDNA yk72e2.3; coded for by C. elegans cDNA yk484a9.3; coded for by C. elegans cDNA yk455d3.3; coded for by C. elegans cDNA yk483g7.3; coded for by C. elegans cDNA yk612d11.3; coded for by C. elegans cDNA yk617c3.3; coded for by C. elegans cDNA yk666e11.3; coded for by C. elegans cDNA yk345h7.3; coded for by C. elegans cDNA yk475d5.3; coded for by C. elegans cDNA yk623g12.3; coded for by C. elegans cDNA yk603f6.3; coded for by C. elegans cDNA yk62a2.3; coded for by C. elegans cDNA yk617c3.5; coded for by C. elegans cDNA yk603f6.5; coded for by C. elegans cDNA yk584g3.5; coded for by C. elegans cDNA yk283h11.5; coded for by C. elegans cDNA yk644h2.5; coded for by C. elegans cDNA yk267e1.5; coded for by C. elegans cDNA yk428c12.5; coded for by C. elegans cDNA yk628e1.5; coded for by C. elegans cDNA yk454d12.5; coded for by C. elegans cDNA yk517d10.5; coded for by C. elegans cDNA yk613d2.5; coded for by C. elegans cDNA yk473f10.5; coded for by C. elegans cDNA yk342b3.5; coded for by C. elegans cDNA yk72e2.5; coded for by C. elegans cDNA yk165h2.5; coded for by C. elegans cDNA yk655a1.5; coded for by C. elegans cDNA yk314d9.5; coded for by C. elegans cDNA yk345h7.5; coded for by C. elegans cDNA yk414a7.5; coded for by C. elegans cDNA yk117f10.5; coded for by C. elegans cDNA yk621e1.5; coded for by C. elegans cDNA yk678f10.5; coded for by C. elegans cDNA yk473f10.3; coded for by C. elegans cDNA yk62a2.5; coded for by C. elegans cDNA yk516d1.5"  
 /codon\_start=1  
 /protein\_id="AA060854.1"  
 /db\_xref="GI:7332167"  
 /translation="MTGSRFTROKVEKPEQVDSDDSDNEAVROELTEORRYAFAG LGDHDQAKQSRSEKAKRLFKSLKQVGYSRICRKSITLPIKPDYKSPG SDYIIIFGEAKIEDLTOHQAOMSAIENLKPFPREPQKTYEEDNEVEDSTGERKD IELVISOANTRRKAIKALKEADNDIVNAIMSLTM"  
 2734..7227  
 /gene="Y65B4BR.4"  
 join(2734..2839,2899..3156,3202..3431,4298..4507,5049..5323,5391..5604,5660..5580,6190..6258,6319..6579,6660..6881,6940..7227)  
 /gene="Y65B4BR.4"  
 /note="contains similarity to Pfam families PF00632 (HECT-domain (ubiquitin-transferase), score=403.2, E=2.5e-117, N=1) and PF00397 (WW domain, score=182.4, E=7.1e-51, E=4); coded for by C. elegans cDNA yk254h3.5; coded for by C. elegans cDNA yk339d4.5; coded for by C. elegans cDNA yk618c2.5; coded for by C. elegans cDNA yk52c9.5; coded for by C. elegans cDNA yk406b11.5; coded for by C. elegans cDNA yk48e8.5; coded for by C. elegans cDNA yk558g7.5; coded for by C. elegans cDNA yk503h4.5; coded for by C. elegans cDNA yk344f2.5; coded for by C. elegans cDNA yk361c11.5; coded for by C. elegans cDNA yk135a2.5; coded for by C. elegans cDNA yk191a2.5; coded for by C. elegans cDNA yk106b12.5; coded for by C. elegans cDNA yk106b12.3; coded for by C. elegans cDNA yk616g10.5"  
 /codon\_start=1  
 /protein\_id="AA060858.1"  
 /db\_xref="GI:7332171"  
 /translation="MARNEPSSQSSGNGTPAOQNGSARFVTVKVNASFYKA ADCYVEITSDPTSSAARKTKYKKTAPAPENNELNHAHESSTISRLLOKALPDPT CIGMAKLKLSLITFRNENGEKFNINISLLAKSSITGINTIFSGYPERKSAVAFETAAASSSPASSTNGVATSSSARAPATKRLTAPATAAANAATGPPACAAE EOLPDGMEKRFDOYGRKYVDHTTKSTTMEPSTOPLPGMEKRRPGRGVYVDINT RTTWOPEFVADMLAEHQWQSGDQAMLQWDELLQNNFSADDDPLGPEWEKRO DPTTSRMKYFVNHNARTTOWEDPTQGRFSDDQLPGMEKRRPTEGQVPEFTIQSKT TYNDPRTGRKAVGLGVGYOMAEKSFPMKIQARFLCSNSNPVHKITVSNNVFE DSFOELMRKNAVDLRRRLVIOERGEIDGCAVARMFLLSHEVLPNPMVCLPMYGN NNSLIQINPASFVNPDIHKTVEEYIGRIFMIRKQGMDDLPKSDOYISSPKRALPHG KPIYSGTTFEYFKMLNKIKVLKLDICQVSEIYNSLMWKDNNIDCDMELRYADIE LIGELTYELKEGTEIAVTEENKLEYIELVEMRNRPNGVEQTKAFPTGFNSVPLE



gene  
CDS  
16304..21657  
/gene="Y65B4BR.3"  
join(16304..16508,17309..17505,18001..18414,18731..19181,  
19416..19990,20620..20693,20741..20927,20978..21361,  
21457..21657)  
/gene="Y65B4BR.3"  
/note="contains similarity of membrane protein  
PATCHED-related proteins"

gene  
CDS  
26298..26394,26997..27159  
/gene="Y65B4BR.2"  
join(21953..22235,24000..24139,25103..25259,25302..25463,  
26298..26394,26997..27159)  
/gene="Y65B4BR.2"  
/codon\_start=1  
/evidence="not\_experimental"  
/product="Hypothetical protein Y65B4BR.2"  
/protein\_id="AAF60856.1"  
/db\_xref="GI:7332169"  
/translation="MENSATISQFTHROKLLALPAFHDLFFPPFPQFLFEISEFPOL  
VECHRLVLFSSAGSPFPPTDLRLVNPVLAFFSILVTOIVPSLDVLSIFSPAGVPRF  
FTFLPVGRRNPVIRLEFNFAACRPKSKPNVDLANLQMDVAVAKAYSLF  
FEMADQIEKFMKQVYVDSKEVKEKCSLEFVDMVLOTPTATFTSKOYGVINDV  
VTNEGYSMGPEBGAVLITGSESDQCPPEPFLGSDGSEQVAYMLSPKRPIM  
VLTRDLTLFEETKREYVDVEKNGFSPMALNRLHFTVDVCRKYNKLXENGAV"  
28359..30706  
/gene="Y65B4BR.1"  
join(28359..28465,28516..28715,28760..28926,28970..29053,  
29103..29176,30205..30706)  
/gene="Y65B4BR.1"

Query Match  
Best Local Similarity 80.0%; Score 16; DB 6; Length 44237;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 aatgacgggaacgtt 17  
|||||  
Db 16259 AATGACGGGAACGTT 16244

RESULT 8  
AC006889/c  
LOCUS  
DEFINITION  
AC006889  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AC006889 267118 bp DNA HTG 26-FEB-1999  
Caenorhabditis elegans clone Y65B4, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
6 unordered pieces.  
AC006889  
HTG: HTGS, PHASE1.  
Caenorhabditis elegans.  
Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
1 (bases 1 to 267118)  
Waterston, R.H.  
The sequence of *Caenorhabditis elegans* clone  
Unpublished  
2 (bases 1 to 267118)  
Waterston, R.H.  
Direct Submission  
Submitted (24-FEB-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

## COMMENT

NOTE: This is a 'working draft' sequence. It currently  
consists of 6 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1  
5875: contig of 5875 bp in length  
5876: gap of unknown length  
5890: contig of 7729 bp in length  
13619: gap of unknown length  
13633: gap of unknown length  
28841: gap of unknown length  
28841: gap of unknown length  
28855: gap of unknown length  
110972: contig of 82117 bp in length  
110985: gap of unknown length  
110986: contig of 63819 bp in length  
174805: gap of unknown length  
174819: contig of 92300 bp in length.

## FEATURES

## source

1..267118  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/clone="Y65B4"

BASE COUNT  
83652 a 30537 c 48664 g 84175 t 70 others  
ORIGIN

Query Match  
Best Local Similarity 80.0%; Score 16; DB 60; Length 267118;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 aatgacgggaacgtt 17  
|||||  
Db 128243 AATGACGGGAACGTT 128228

RESULT 9  
AE005077  
LOCUS  
DEFINITION  
AC005077 AE004437  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AE005077 14625 bp DNA BCT 12-FEB-2001  
Halobacterium sp. NRC-1 section 108 of 170 of the complete genome.  
AE005077.1 GI:10581157  
Halobacterium sp. NRC-1.  
Halobacterium sp. NRC-1.  
Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
Halobacterium.

## REFERENCE

## AUTHORS

1 (bases 1 to 14625)  
Ng, M.V., Kennedy, S.P., Mahairas, G.G., Bergquist, B., Pan, M.,  
Shukla, H.D., Lasky, S.R., Baliga, N., Thorsson, V., Shrogha, J.,  
Swartzell, S., Weir, D., Hall, J., Dahl, T.A., Weir, R., Goo, Y.A.,  
Leitner, B., Keller, K., Cruz, R., Danson, M.J., Hough, D.W.,  
Maddocks, D.G., Jablonski, P.E., Krebs, M.P., Angelvine, C.M., Dale, H.,  
Isenbarger, T.A., Peck, R.F., Pohlisch, M., Spudis, J.L.,  
Jung, K.-H., Alam, M., Freitas, T., Hou, S., Daniels, C.J., Dennis, P.P.,  
Omer, A.D., Edhardt, H., Lowe, T.M., Llang, P., Riley, M., Hood, L. and  
DasSarma, S.  
From the cover: genome sequence of *Halobacterium* species NRC-1  
Proc. Natl. Acad. Sci. USA 97 (22), 12176-12181 (2000)  
11016950  
2 (bases 1 to 14625)

AUTHORS Ng,W.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M., Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,V., Sbrogna,J., Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Meli,R., Goo,Y.A., Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W., Meddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H., Isenbarger,T.A., Peck,R.F., Ponschood,M., Spudich,U.L., Jung,K.H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P., Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and Dassarma,S.

Direct Submission  
Submitted (14-JUL-2000) Institute for Systems Biology, 4225  
Roosevelt Way NE, Seattle, WA 98105, USA

FEATURES  
Location/Qualifiers  
1. .14625  
/organism="Halobacterium sp. NRC-1"  
/strain="NRC-1"  
/db\_xref="taxon:64091"  
complement(83. .153)  
/gene="trn29"  
/product="tRNA-Gly"  
complement(83. .153)  
/gene="trn29"  
401. .1252  
/gene="VNG1688C"  
401. .1252  
/gene="VNG1688C"  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="Vng1688C"  
/protein\_id="AAG19935.1"  
/db\_xref="GI:10581158"  
/translation="MTRSVLVPSLSLVREAEKKREATKLGVAARAAVFRIDRVVPP  
DEDEGRQMGGEFETVLRATPPYLKKEAFDRDELAYGLPLRLSSMTGSDSSG  
SGSLRQGIIVTVGSEGRVAGCMQHPISLRHGGMAVSEGERVITRVSRRPYRAKL  
VDPILPGEVSVERTGIDGALDRSDAGVRIATSRHGGPELSVSLGCRYRIRACDVTAF  
GABERCLPPMLGVASDAVNESVTDSSADAPARDAWLNTIPDQSEVYRTAEVLAATL  
GSILFTE"  
1257. .2264  
/gene="trp13p"  
/note="VNG1689G"  
1257. .2264  
/gene="trp13p"  
/note="Rpl3p"  
/codon\_start=1  
/transl\_table=11  
/product="50S ribosomal protein L13p"  
/protein\_id="AAG19936.1"  
/db\_xref="GI:10581159"  
/translation="MPQPNRPKRGSMGFSFRKRAESEVPRFNSMPADGVEGLQGFAG  
YKAGMTHVLVDKNAKPTGMEITYVTVETPPMRAAVRLIEDPYGKKPLTEVM  
ADPTHESLDRTLSPDEGSDTDELEALDTEELADIIVITHTVPGDAGVKKPLTEVM  
ETVYGGGTADRLLEFADLIEDGVHAFDGFPRAGEFTDAAGITKGGTQGPVKRMVY  
QKRGKARQGRGRRRIGNGIGMNPISRVSRYVPOOGSGYQHORTEINKRLDIDNGDP  
TPDGGPRNYGEVDPPTLVKGSVPGPEQRLVRRPRANRPNESRLDPEVRYVSTASNQ  
G"  
2268. .3020  
/gene="rpl4e"  
/note="VNG1690G"  
2268. .3020  
/gene="rpl4e"  
/note="Rpl4e"  
/codon\_start=1  
/transl\_table=11  
/product="50S ribosomal protein L4E"  
/protein\_id="AAG19937.1"  
/db\_xref="GI:10581160"  
/translation="MOTVTRDLDDAGTLDLPVREEPPRPDLVKRAVLAQAARNTQ  
EYGADEAYAGRTAESQSGSGMAHYKANGOGARVPQTVGGRKAHPKRAEKHGLDV  
NDEKRAVRAAAATDSELVADRGHNFDDVDEPLVVSDDFEDLVKTDVVSLLEA  
LGYHADIERADEGRITVAGOGTLRGKRYOEPTSLIVTASESGPSTRAARLADYVAT  
GREVNADDLAPGAEPPGRITVWTSAAVEEVAQR"  
3017. .3271  
gene

CDS  
/gene="rpl23p"  
/note="VNG1691G"  
3017. .3271  
/gene="rpl23p"  
/note="Rpl23p"  
/codon\_start=1  
/transl\_table=11  
/product="50S ribosomal protein L23p"  
/protein\_id="AAG19938.1"  
/db\_xref="GI:10581161"  
/translation="MSGIIDYPLVTEKAMDEMPQNKQPIFYIDAAKPEIRDVSESE  
IDVTVVDVNTQITPEAKKATVKLASDDQDVASRLGVF"  
3274. .3996  
/gene="rpl2p"  
/note="VNG1692G"  
3274. .3996  
/gene="rpl2p"  
/note="Rpl2p"  
/codon\_start=1  
/transl\_table=11  
/product="50S ribosomal protein L2p"  
/protein\_id="AAG19939.1"  
/db\_xref="GI:10581162"  
/translation="MGRRIQGRQGRGTSTFRAPSHRYKALSHKRTEDTVLAGVI  
DVEHDPARSAFVARVAFEDDQRLVLASEVGVDTEIGISATIEBGNPLAETPE  
GVPCVNEVSHPDGGRFARAGVNAADLVYTERATIVLEPGSETKRLSPDCRATIGV  
AGGRTKEPVPKAGNKKHKKMARGCTWPRVGVAMNAVDPFGGGGRQHPGRKSVSR  
DAAPGRKVDIASKRTGCGNE"  
4000. .4422  
/gene="rps19p"  
/note="VNG1693G"  
4000. .4422  
/gene="rps19p"  
/note="Rps19p"  
/codon\_start=1  
/transl\_table=11  
/product="30S ribosomal protein S19p"  
/protein\_id="AAG19940.1"  
/db\_xref="GI:10581163"  
/translation="MSPEYRTGRGREGFTYRGHDLDELQEMSLBDVAELLPARQRTIT  
RGLSEHHKVLAEARSGTEETANNPTRTHLRQMPVLPFVGLTPAVYTGQEEEREVEY  
QPMIGHYLGEPQLTRSSVEHQAGIGATSSKFPVLK"  
4426. .4896  
/gene="rpl22p"  
/note="VNG1695G"  
4426. .4896  
/gene="rpl22p"  
/note="Rpl22p"  
/codon\_start=1  
/transl\_table=11  
/product="50S ribosomal protein L22p"  
/protein\_id="AAG19941.1"  
/db\_xref="GI:10581164"  
/translation="MGISYVDVDSASAKMLRERSISLKHKAIRAEISGFTVADA  
KEYLQAVTIDERSVPFRKQNSGYGHRNDIDGMRAGRIPKASADFLKLNSVNNADQ  
QGFDADEMTVEHVAHVKVSGSQRKPRAMGRATVWMTLCDVEIVTEEVTA"  
4896. .5804  
/gene="rps3p"  
/note="VNG1697G"  
4896. .5804  
/gene="rps3p"  
/note="Rps3p"  
/codon\_start=1  
/transl\_table=11  
/product="30S ribosomal protein S3p"  
/protein\_id="AAG19942.1"  
/db\_xref="GI:10581165"  
/translation="MADELEFIEQGIQSRQIDEFFAEELARAGYGGNELAPTPMGNOI  
VLAEKRPQWVIGSGKQIRKITTOLEERFPLEDPDQVDOVEEPDILNAQIVAPRLANA  
LERGAPFRKAGHTTIRIMSGALAEIYVLSGTPQNRGKVEKFNNGYIKHNEPPEE  
IYHKGAVAMKIGTIGVNYKTIIPNAELPDPEIQDADIEDLVYDEAFAGAGDLEL  
LEGEDADADADADAAEPESEPADEFEEVETIDDDVEBELDELADAVEGEDEDEDF  
SDVDEADAPYADBLDMDDEDGAE"

gene 5804. .6019  
/gene="rp129p"  
/note="VNG1698G"  
5804. .6019  
/gene="rp129p"  
/note="Rp129p"  
/codon\_start=1  
/transl\_table=1  
/product="50S ribosomal protein L29p"  
/protein\_id="AAG19943.1"  
/db\_xref="GI:10581166"  
/translation="MAILVTEIRDMTPAEREAELELRTELNSKAVKAAGAPDNP  
GRISLKKTIARKTAVOREGDLADDE"  
6028. .6339  
/gene="VNG1699C"  
6028. .6339  
/gene="VNG1699C"  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=1  
/product="Vng1699C"  
/protein\_id="AAG19944.1"  
/db\_xref="GI:10581167"  
/translation="MTTTPESLPRLHGLHARVADTDGSRVGIAGRVVDETMTQV  
VLRTASGVAVPKKATFERLTDDEAAPPNGVGTAKPAGGETROTGESVAVTVD  
GGRLSRPERSENGVDSKMR"  
6390. .6719  
/gene="rps17p"  
/note="VNG1700G"  
6390. .6719  
/gene="rps17p"

Query Match 79.0%; Score 15.8; DB 1; Length 14625;  
Best Local Similarity 89.5%; Pred. NO. 8.2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 caatgacggaacgctcc 19  
||||| ||||| |||  
Db 8856 CAATGACGGAACGATCC 8874

RESULT 10  
LOCUS MTW006 22440 bp DNA BCT 18-JUN-1998  
DEFINITION Mycobacterium tuberculosis H37Rv complete genome: segment 54/162.  
ACCESSION AL021006 AL123456  
VERSION AL021006.1 GI:3242277  
KEYWORDS  
SOURCE Mycobacterium tuberculosis.  
ORGANISM Mycobacterium tuberculosis  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;  
Mycobacterium.

REFERENCE 1 (bases 1 to 22440)  
AUTHORS Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,  
Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,  
Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,  
Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,  
Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,  
Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,  
Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,  
Squares,S., Squares,R., Sultison,J.E., Taylor,K., Whitehead,S. and  
Barrell,B.G.

TITLE Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence  
JOURNAL Nature 393 (6685), 537-544 (1998)  
MEDLINE 98295987  
REMARK Erratum: [[Published erratum appears in Nature 1998 NOV  
12:396(6707):190]]  
2 (bases 1 to 22440)  
AUTHORS Parkhill,J.  
TITLE Direct Submission  
JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium

COMMENT  
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome  
Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique  
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,  
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
On Jun 20, 1998 this sequence version replaced gi:2695814.  
Notes:  
Details of M. tuberculosis sequencing at the Sanger Centre are  
available on the World Wide Web.  
(URL, [http://www.sanger.ac.uk/Projects/M\\_tuberculosis/](http://www.sanger.ac.uk/Projects/M_tuberculosis/)) CDS have  
been renumbered from the original cosmid submissions but the old  
gene designations are in brackets after the new gene numbers.  
Gene prediction was based on a Hidden Markov Model of TB genes  
implemented in Tbpase (Krogh) supplemented with visual inspection  
of positional base preference in codons, especially where there is  
an increase in the observed/expected third position G + C.  
CAUTION: In some cases we may not have predicted the correct  
initiation codon. Where possible we choose an initiation codon  
(atg, gtg, or ttg) which is preceded by an upstream ribosome  
binding site sequence (optimally 5-13bp before the initiation  
codon). If this cannot be identified we choose the most upstream  
initiation codon.

FEATURES  
source  
location/Qualifiers  
1. .22440  
/organism="Mycobacterium tuberculosis"  
/strain="H37Rv"  
/db\_xref="taxon:1773"  
<1. .>22440  
/note="fragment designated v006. Does not represent a  
physical clone"  
108. .112  
/note="possible RBS, agagg, for RV1234"  
124. .651  
/gene="RV1234"  
124. .651  
/note="RV1234"  
/note="RV1234", (MTW006.06), len: 175. Unknown. Tbpase  
score is 0.929"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein RV1234"  
/protein\_id="CAI15890.1"  
/db\_xref="GI:2695820"  
/db\_xref="SPTREMBL:O50451"  
/translation="MTSPFOPRQVPGSTPAAGAGRGVPAIPPPKGPVGYPTVA  
EAORAVDLSQDFPVQOVTIVGDLMQVEVETRLTMPKVLGGVSGALGIFIGL  
VIGFSPMPMSALVTGLVAGVFPLITISAVPYANARGRDPSSTMTQLVAGRYDLCP  
ONAEKARDLARLAI"  
663. .666  
/note="possible RBS, gagg, for RV1235"  
672. .2078  
/gene="1pgy"  
672. .2078  
/note="1pgy"  
/gene="RV1235", (MTW006.07), len: 468. Unknown 11poprotein  
highly similar to M. leprae protein u1756v (469 aa),  
possibly involved in sugar (e.g. maltose) uptake. FASTA  
scores: MU01518034opt: 2442 E(): 0: (77.4% identity in 470  
aa overlap). Similar to P18815(MAL:ENTAE MALTOSE-BINDING  
PERIPLASMICPROT (396 aa)FASTA scores: opt: 193 E():  
2.3e-05; (24.2% identity in 297 aa overlap). Contains  
PS00013 Prokaryoticmembrane11poprotein 11lipid attachment  
site. Tbpase score 150.914"  
/codon\_start=1  
/transl\_table=1  
/product="1pgy"  
/protein\_id="CAI15891.1"  
/db\_xref="GI:2695821"  
/db\_xref="SPTREMBL:O1095"  
/translation="TMSRGRIPRLGAVALVLTAAACGADSGGLVSEFYTPATDG  
ATPFAIORCNOGCGRGTAAOVSLPSPNROQLARLTGNDRTDVALVWTA  
EFAEGVALPLSDPDAELANDAVADLPPLATAGNNHKLTYAPVYTNQOLLYPRD  
LVNSPPTDWMAMIAEARLHAAGEPSIAVQANGDEGLVWVFNLVASGASVLSDEG  
RHWVLTDTPAHRAVTSALQILKSVATTPGADPSITREBGSARLAEQKAALEVNM

```

misc_feature
/feature="lpgy"
/note="P500013 Prokaryotic membrane lipoprotein lipid
attachment site"
2075..2998
/feature="suga"
2075..2998
/feature="suga"
/feature="suga"
/note="Rv1236, (MTV006.08), len: 307. Unknown membrane
protein probably involved in sugar transport, most similar
to M. leprae protein designated malg, U15180|MLU1518035
(310 aa). Also similar to numerous bacterial sugar
transport system components. FASTA scores: opt: 1566 E():
0; (81.8% identity in 292aa overlap). Contains P500402
Blinding-protein-dependent transport systems inner membrane
comp signature. Also similar to MTCY3612.18c, (290 aa) opt:
514, E(): 7.3e-27; (33.2% identity in 283 aa overlap) and
MTCY1687.08. Tbpasescore is 0.899"
/codon_start=1
/transl_table=11
/product="suga"
/protein_id="CA15892.1"
/db_xref="GI:2695822"
/db_xref="SPTREMBL:O50452"
/translation="MTSEBORATAVESRTGSMARRRLAFMLVAPAAMLAVTAVP
IGALMLSLORNLATPNDATFICGNHYTHILLDRWMALATLATITANSVTEVL
GALALVYHRTILIGKGLVTRAVLIPGITYVVASYSMTYAMTGTGLANLPLDYSAP
LTQQLPSLGLVIAEYVKTTPKPSLLGLALVAPEDLLRAAVDASAWRLTKYL
PMKPAIVALLFRTLDARIFDNIYVLVGLGSGNMTGSVSLIGDNLFKGNVGLGSAI
SVLIFGCAVIAFRTIFKLGAAAPGSGPSGR"
2645..2731
/feature="suga"
/note="P500402 Binding-protein-dependent transport
systems inner membrane comp signature"
3003..3827
/feature="sugB"
3003..3827
/feature="sugB"
/note="Rv1237, (MTV006.09), len: 274. Probable sugar
(e.g. maltose) transport protein similar to numerous
bacterial sugar transport protein. Most similar to M.
leprae protein designated Malg (296 aa). FASTA scores:
U15180|MLU1518036opt:1571E(): 0; (89.8% identity in 274 aa
overlap). Related to MTCY1687.08, (275 aa) FASTA
scores:opt: 370, E(): 2.4e-17; (26.8% identity in 269 aa
overlap). Tbpase score 150.895"
/codon_start=1
/transl_table=11
/product="sugB"
/protein_id="CA15893.1"
/db_xref="GI:2695823"
/db_xref="SPTREMBL:O50453"
/translation="MGARATYAWADVLTGVALPLVIMFSLKPTSTGVKILI
PSTVEDNRGIFRGDLFSSALNSIGITIVIVAGAAVAVARLEPFGKRL
IGALLITMPEPSISIVPLPENTERAIGLPTGCLIPPLATITLSPFEI
PMDLEKAKMDGRTPGQAFRRKYVPLAAGVTAALVLFPMNDLILSTRATRA
ITRAVAIANFTGSQFEPTGSIAGAIVITPIVFLFORIVAGLTSVAVK"
3832..5013
/feature="sugC"
3832..5013
/feature="sugC"
/note="Rv1238, (MTV006.10), len: 393. Unknown ABC
transporter component of sugar uptake system. Highly
similar to M. leprae protein designated ugpc (392 aa) in
U15180. FASTA score: opt:2007 E(): 0; (79.9% identity in
389 aa overlap). Contains P500017 ATP/GTP-binding site
motif A (P-loop) and P500211 ABC transporters family
signature. Tbpase score is 0.920"
/codon_start=1

```

```

Query Match 79.0%; Score 15.8; DB 3; Length 22440;
Best Local Similarity 89.5%; Pred. No. 8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 caatgacgggaacgttcc 19
Db 15583 CAAAGCAGGGAAGCGTCC 15565

RESULT 11
CELFA4E7/C 22480 bp DNA INV 08-JUL-1998
LOCUS
DEFINITION
Caenorhabditis elegans cosmid F44E7.
ACCESSION
AF016421
VERSION
AF016421.1 GI:3294490
KEYWORDS
SOURCE
Caenorhabditis elegans.
ORGANISM
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
1 (bases 1 to 22480)
WILSON, R., AINSWORTH, R., ANDERSON, K., BAYNES, C., BERTS, M.,
BOFIELD, J., BURTON, J., CONNELL, M., COPSEY, T., COOPER, J.,
COULSON, A., CRAXTON, M., DEAR, S., DU, Z., DURBIN, R., FAVELLO, A.,
FULLON, L., GARDNER, A., GREEN, P., HAWKINS, T., HILLIER, L., JIER, M.,
JOHNSTON, L., JONES, M., KERSHAW, J., KIRSTEN, J., LAISTER, N.,
LATREILLE, P., LIGHTNING, J., LLOYD, C., MCMURRAY, A., MORIMORE, B.,
O'CALLAGHAN, M., PARSONS, J., PERRY, C., RIFKEN, L., ROOPER, A.,
SAUNDERS, D., SHOWNKEEN, R., SMALDON, N., SMITH, A., SONNHAMMER, E.,
STAUDEN, R., SULISTON, R., THIERLY-MIEG, J., THOMAS, K., VAUDIN, M.,
VAUGHAN, K., WATERSTON, R., WATSON, A., WEINSTOCK, L.,
WILKINSON-SPOAT, J. and WOHLDMAN, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE
2 (bases 1 to 22480)
Du, Z., Scheet, P. and Delehaunty, A.
The sequence of C. elegans cosmid F44E7
JOURNAL Unpublished (1998)
REFERENCE
3 (bases 1 to 22480)
Waterston, R.
Direct Submission
JOURNAL Submitted (29-JUL-1997)
REFERENCE
4 (bases 1 to 22480)
Waterston, R.
Direct Submission
JOURNAL Submitted (08-JUL-1998) Department of Genetics, Washington

```

## COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Jul 8, 1998 this sequence version replaced gi:2291178.  
Submitted by:

Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
e-mail: rwenematode.wustl.edu and jesesanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate  
chemistry; an attempt was made to resolve all sequencing problems,  
such as compressions and repeats; all regions were covered by  
sequence from more than one subclone

## NEIGHBORING COSMID INFORMATION:

The 5' cosmid is C53A3. 700 bp overlap; 3' cosmid is ZC250, 200 bp  
overlap. Actual start of this cosmid is at base position 501 of  
CELF44E7; actual end is at 16254 of CELC250

## NOTES:

Coding sequences below are predicted from computer analysis, using  
the program GeneFinderP. Green and L. Hillier, ms in preparation).

## Location/Qualifiers

1. 22480

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db\_xref="taxon:6239"

/clone="F44E7"

/chromosome="V"

255. 1363

/gene="F44E7.2"

/join(255. 417,468. 1140,1192. 1363)

/gene="F44E7.2"

/note="Similar to nltrophenyolphosphatase: coded for by C.  
elegans cDNA yk312h2.5; coded for by C. elegans cDNA  
yk267e8.5; coded for by C. elegans cDNA yk312h2.3"

/codon\_start=1

/protein\_id="AAC25793.1"

/db\_xref="GI:3294491"

/translation="MKYPTIDSOQLRFRPHPLKAGLDPKCRSTKKPLCDPTPAKVK  
TIDFTIDADGVLMGESSVMPGSPRLIDYLVKHKQITVLNNATKRAVYAKKLAKV  
GYNSSKMNKNLVNPAVAVADTLRAGLDGRVLLIGQGLRDEMDLGIEYFGGPE  
KKQEDAGSGAFMYDILKENVGAVVGEYKHPYIKMKASNYLRGCVLFTNED  
ETGCPNPEVYIPAGPIVAIKCAGSDPLTGVKPCPAVNYIKRKINNPSTMTI  
GDRNTVVKFGROHGMKTLVLVSGCHQIEDIIEQNMRDMVDPYVAPCGALVPER  
I"

2001. 2253

/gene="F44E7.3"

/join(2001. 2045,2095. 2253)

/gene="F44E7.3"

/codon\_start=1

/evidence="not experimental"

/protein\_id="AAC25788.1"

/db\_xref="GI:2291180"

/translation="MTLQLOELMRQSDKEIEVLELQSRLELAQAASDLTDRLEHDLRA  
SDAVQETFKFLTPQPRNOKME"

complement(2795. 6738)

/gene="F44E7.4"

complement(join(2795. 3047,3095. 3431,3478. 4078,  
4348. 4482,4530. 5338,5385. 5652,5685. 6253,6307. 6451,  
6632. 6738))

gene

CDS

CDS

/gene="F44E7.4"

/note="strong similarity to peptidase family M16 or the  
insulinase family: coded for by C. elegans cDNA yk18e4.3;  
coded for by C. elegans cDNA yk6c6.3; coded for by C.  
elegans cDNA yk87d9.3; coded for by C. elegans cDNA  
yk133a3.3; coded for by C. elegans cDNA yk34h4.3; coded  
for by C. elegans cDNA yk15f11.3; coded for by C. elegans  
cDNA yk5a5.5; coded for by C. elegans cDNA yk182f7.5;  
coded for by C. elegans cDNA yk87g9.5; coded for by C.  
elegans cDNA yk133a3.5; coded for by C. elegans cDNA  
yk18e4.5; coded for by C. elegans cDNA yk15e2.5; coded  
for by C. elegans cDNA yk178h6.5; coded for by C. elegans  
cDNA yk6e1.5; coded for by C. elegans cDNA yk179h10.5;  
coded for by C. elegans cDNA yk34h4.5; coded for by C.  
elegans cDNA yk6c6.5"

/codon\_start=1

/protein\_id="AAC25789.1"

/db\_xref="GI:2291181"

/translation="MSLSLSKFGKQVSPALAHIAKRTSSVNSQILRRPQNSLR  
LRIVNINSNMPLPKMTAGKNIVLAKRDLIVKAQDAREYRGLETGIVLVSDP  
TTDKSAALDVKGVGLMPWELPGIAHCEHMLFLGTAKKYPSENEYSFLAAGSSN  
AYTSDHTNHYEDVKPPDLPALDRFVQFSLPOTGESATREYCAVDSNNLND  
LMRLQVDRSRKPGHDIKFTGKQTLLEDARKKIEPPDALLQHKKYSDDINT  
CCIVGKEPLNVLESYLTLEFDALENKRVKWEPEYGPDLAKRIDVVPIDTRL  
VSISFEPDLNGEFLSQPGHYISHLIGHEGCSLSLELRUGVSSLSQSDHTQAG  
GVYNTMDLSTEGEIVDEITQLMFNYIGMLOSAGPKQWVDELAELSAVFRKDE  
OPMTAIVASAIQYIPREHILSSRYLTITKYEPRIRKLSMLSPANVQVRSOKK  
GQEGNTNRPVYGTETKATYDISPEMKKTKENLAKTSHALHLPENREYATNPDKPRE  
SVNKHPRILISDDGKRWYRKODDETNPQKOTELATLTPVAVQNPRLNSLMLWC  
LSDLTAEETVNAADLAGLKCQLESSPFGVQMVSNRRREARHASTLVHAYGDEQALP  
AKHLANRMTNFKIDKTRDVLFEESLKRATLNHASFQPYLLOHYNOLIVDKVSKBO  
LLAVCDSTLEEDVOGFAKEMQAEMLEFVNGSTEKALQISELMQVLSKAPNSR  
PLVNEHNPRELONNGDEVYVRLQTHVGCVEVYQIGVONTYDNVAVGLIDOL  
IREPAFNTLRNEALGYIVMGVSRILNGSTVALNTVQSPKSDVHLREIEVLSVK  
ELAEPRQEPNQVSGMALARLEKPKTILSSFRFRPMNIECRQVNFARREEVALLKT  
IKDDVLELFDKTKIRKDAERKRLAVFHGNEQEAANTTIKNAESGKKEKYLVS  
DQLRQFLPLYKRIADLKPIDVPLEHOETTSKY"

complement(8821. 10572)

/gene="F44E7.5"

complement(join(8821. 9324,9577. 9913,10129. 10244,  
10441. 10572))

/gene="F44E7.5"

/codon\_start=1

/evidence="not experimental"

/protein\_id="AAC25790.1"

/db\_xref="GI:2291182"

/translation="WELSTDLMKETSISLADHPSPCLLSSSSKRLRETNQVIGHYE  
AHLHITLPEKDCDEKCKSKERQFEMILNKKSGTRRRAHFHSIFCLDIDLAYN  
DLLEFAKELITLRNPNPENSDFVFNKHSLEKCEDICFNGSLPVHNGVSMFHSC  
HNMRKLVISNPTDTSFQFMPKLGSEVFETDLMIHSEVDLEAINSLKYSRLKSYE  
FCFSKATIDYATAILOKCDQVKCFENFGHSGCVAPVIRLLTDIPRSIVFGCVA  
CRSEGTPEVGGFMENGVNAVVOYVTKLLERDRFSLVGKTLADAVLEFRPNANSVS  
MYIQKDSVCGPNCLDIEHSQKAGVL"

complement(11113. 12353)

/gene="F44E7.6"

complement(join(11113. 11319,11364. 11756,11874. 11997,  
12043. 12129,12177. 12355))

/gene="F44E7.6"

/codon\_start=1

/evidence="not experimental"

/protein\_id="AAC25791.1"

/db\_xref="GI:2291183"

/translation="MDPRALRICVLFQHLQSGIEFNVAVKAMSKVFGDHMEYRHFEVY  
YMGANGSLGTVDSYDRHTSERKFEFTGVQVAMITDKLNPIDLEFARAACRQPKDI  
IDNOKSTLAKTIMNIDETRVDELNNNSIRCLINDCEPENVIMMMDEKVRLEFLAIKQI  
LESTIGLSAKSVLINMDQTYEIAELQHVAKREVLEDLITKLGSERENVLGEPMKKA  
KNLKLAKVPHSITIVKHLNHRSMNPFVSGSFRRSDITKRDVELLTKSTNFEVFTSE  
NVKFSARLIVVPIHLSPPSSSGSIVYKSKGTIDYVHPNGFTIKKKIDSNH"

complement(12689. 15444)

/gene="F44E7.7"

complement(join(12689. 12910,12966. 13058,13111. 13179,  
13726. 14592,14894. 15112,15159. 15222,15357. 15444))

/gene="F44E7.7"

## FEATURES

source

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS





TITLE  
JOURNAL

## COMMENT

Macdonald, P., Marguis, N., McEwan, P., McGurk, A., McKernan, K.,  
McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J.,  
Norman, C. H., O'Connor, T., O'Donnell, P., Olivier, T. M., Peterson, K.,  
Pierre, N., Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,  
Zimmer, A. and Zody, M.  
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 13, 2000 this sequence version replaced 91:6705581.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: L5496

Center clone name: 319\_D\_2

NOTE: This record contains 75 individual  
sequencing reads that have not been assembled into  
contigs. Runs of N are used to separate the reads  
and the order in which they appear is completely  
arbitrary. Low-pass sequence sampling is useful for  
identifying clones that may be gene-rich and allows  
overlap relationships among clones to be deduced.  
However, it should not be assumed that this clone  
will be sequenced to completion. In the event that  
the record is updated, the accession number will  
be preserved.

1  
844 943: gap of 100 bp in length  
944 1807: contig of 864 bp in length  
1808 1907: gap of 100 bp  
1908 2770: contig of 863 bp in length  
2771 2870: gap of 100 bp  
2871 3770: contig of 900 bp in length  
3771 3870: gap of 100 bp  
3871 4722: contig of 852 bp in length  
4723 4822: gap of 100 bp  
4823 5702: contig of 880 bp in length  
5703 5802: gap of 100 bp  
5803 6667: contig of 865 bp in length  
6668 6767: gap of 100 bp  
6768 7657: contig of 890 bp in length  
7658 7757: gap of 100 bp  
7758 8637: contig of 880 bp in length  
8638 8737: gap of 100 bp  
8738 9612: contig of 875 bp in length  
9613 9712: gap of 100 bp  
9713 10583: contig of 871 bp in length  
10584 10683: gap of 100 bp  
10684 11555: contig of 872 bp in length  
11556 11655: gap of 100 bp  
11656 12512: contig of 857 bp in length  
12513 12612: gap of 100 bp  
12613 13467: contig of 875 bp in length  
13468 13587: gap of 100 bp  
13588 14468: contig of 881 bp in length  
14469 14568: gap of 100 bp  
14569 15428: contig of 860 bp in length  
15429 15528: gap of 100 bp  
15529 16398: contig of 870 bp in length  
16399 16498: gap of 100 bp  
16499 17379: contig of 881 bp in length  
17380 17479: gap of 100 bp  
17480 18364: contig of 885 bp in length  
18365 18464: gap of 100 bp

18465 19354: contig of 890 bp in length  
19355 19454: gap of 100 bp  
19455 20340: contig of 886 bp in length  
20341 20440: gap of 100 bp  
20441 21310: contig of 870 bp in length  
21311 21410: gap of 100 bp  
21411 22303: contig of 893 bp in length  
22304 22403: gap of 100 bp  
22404 23256: contig of 853 bp in length  
23257 23356: gap of 100 bp  
23357 24243: contig of 887 bp in length  
24244 24343: gap of 100 bp  
24344 25189: contig of 846 bp in length  
25190 25289: gap of 100 bp  
25290 26151: contig of 862 bp in length  
26152 26251: gap of 100 bp  
26252 27073: contig of 822 bp in length  
27074 27173: gap of 100 bp  
27174 28052: contig of 879 bp in length  
28053 28152: gap of 100 bp  
28153 29054: contig of 902 bp in length  
29055 29154: gap of 100 bp  
29155 29964: contig of 810 bp in length  
29965 30064: gap of 100 bp  
30065 30946: contig of 882 bp in length  
30947 31046: gap of 100 bp  
31047 31926: contig of 880 bp in length  
31927 32026: gap of 100 bp  
32027 32901: contig of 875 bp in length  
32902 33001: gap of 100 bp  
33002 33840: contig of 839 bp in length  
33841 33940: gap of 100 bp  
33941 34821: contig of 881 bp in length  
34822 34921: gap of 100 bp  
34922 35795: contig of 874 bp in length  
35796 35895: gap of 100 bp  
35896 36723: contig of 828 bp in length  
36724 36823: gap of 100 bp  
36824 37709: contig of 886 bp in length  
37710 37809: gap of 100 bp  
37810 38681: contig of 872 bp in length  
38682 38781: gap of 100 bp  
38782 39644: contig of 863 bp in length  
39645 39744: gap of 100 bp  
39745 40630: contig of 886 bp in length  
40631 40730: gap of 100 bp  
40731 41580: contig of 850 bp in length  
41581 41680: gap of 100 bp  
41681 42555: contig of 875 bp in length  
42556 42655: gap of 100 bp  
42656 43471: contig of 816 bp in length  
43472 43571: gap of 100 bp  
43572 44433: contig of 862 bp in length  
44434 44533: gap of 100 bp  
44534 45401: contig of 868 bp in length  
45402 45501: gap of 100 bp  
45502 46365: contig of 864 bp in length  
46366 46465: gap of 100 bp  
46466 47335: contig of 870 bp in length  
47336 47435: gap of 100 bp  
47436 48286: contig of 851 bp in length  
48287 48386: gap of 100 bp  
48387 49278: contig of 892 bp in length  
49279 49378: gap of 100 bp  
49379 50265: contig of 887 bp in length  
50266 50365: gap of 100 bp  
50366 51235: contig of 870 bp in length  
51236 51335: gap of 100 bp  
51336 52218: contig of 883 bp in length  
52219 52318: gap of 100 bp  
52319 53193: contig of 875 bp in length  
53194 53293: gap of 100 bp  
53294 54147: contig of 854 bp in length



```
* 54148 54247: gap of 100 bp
* 54248 55137: contig of 890 bp in length
* 55138 55237: gap of 100 bp
* 55238 56111: contig of 874 bp in length
* 56112 56211: gap of 100 bp
* 56212 57071: contig of 860 bp in length
* 57072 57171: gap of 100 bp
* 57172 58030: contig of 859 bp in length
* 58031 58130: gap of 100 bp
* 58131 59013: contig of 883 bp in length
* 59014 59113: gap of 100 bp
* 59114 59982: contig of 869 bp in length
* 59983 60082: gap of 100 bp
* 60083 60966: contig of 884 bp in length
* 60967 61066: gap of 100 bp
* 61067 61935: contig of 869 bp in length
* 61936 62035: gap of 100 bp
* 62036 62909: contig of 874 bp in length
* 62910 63009: gap of 100 bp
* 63010 63816: contig of 807 bp in length
* 63817 63916: gap of 100 bp
* 63917 64804: contig of 888 bp in length
* 64805 64904: gap of 100 bp
* 64905 65793: contig of 891 bp in length
* 65796 65895: gap of 100 bp
* 65896 66766: contig of 871 bp in length
* 66767 66866: gap of 100 bp
* 66867 67702: contig of 836 bp in length
* 67703 67802: gap of 100 bp
* 67803 68676: contig of 874 bp in length
```

```
Query Match 79.0%; Score 15.8; DB 67; Length 72593;
Best Local Similarity 89.5%; Pred. No. 7.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 1 caatgacgggaaagcttc 19
    ||||| ||||| |||||
Db 66710 CAATGTCGGGAATCGTCC 66692
```

RESULT 15  
LMFLCHR34\_1/c  
WPCOMMENT

Sequence split into 7 fragments LOCUS LMFLCHR34 Accession AL499623

Fragment Name	Begin	End
LMFLCHR34_0	1	110000
LMFLCHR34_1	100001	210000
LMFLCHR34_2	200001	310000
LMFLCHR34_3	300001	410000
LMFLCHR34_4	400001	510000
LMFLCHR34_5	500001	610000
LMFLCHR34_6	600001	697361

Continuation (2 of 7) of LMFLCHR34 from base 100001 (AL499623 *Leishmania* major chromosome)

```
Query Match 79.0%; Score 15.8; DB 84; Length 110000;
Best Local Similarity 89.5%; Pred. No. 7.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 1 caatgacgggaaagcttc 19
    ||| ||||| ||||| |||||
Db 46221 CAACGACGCGGAGACGTCC 46203
```

Search completed: June 8, 2001, 06:03:00  
Job time: 7310 sec

• • • • •

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2001, 06:21:25 ; Search time 217.33 Seconds  
(without alignments)  
53.723 Million cell updates/sec

Title: US-09-507-242-6  
Perfect score: 20  
Sequence: 1 caatgacgggaacgttcg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 678276 seqs, 291890651 residues  
Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_0401.\*  
1: /SID56/gcgdata/geneseq/geneseqn/NA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseqn/NA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseqn/NA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseqn/NA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseqn/NA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseqn/NA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseqn/NA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseqn/NA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseqn/NA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseqn/NA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseqn/NA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseqn/NA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseqn/NA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseqn/NA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseqn/NA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseqn/NA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseqn/NA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseqn/NA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseqn/NA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT.\*  
22: /SID56/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	A50294	Candida albicans C
2	20	100.0	989	A50292	Candida albicans C
3	15.2	76.0	274	T24451	Human gene signatu
4	15.2	76.0	479	Z91619	E.coli promoter re
5	15.2	76.0	560	C04556	Human secreted pro
6	15.2	76.0	603	C42788	Arabidopsis thalia
7	15.2	76.0	1062	C48389	Arabidopsis thalia
8	15.2	76.0	1315	F13745	Aspergillus oryzae
9	15.2	76.0	1473	X61690	B. burgdorferi ant
10	15.2	76.0	1527	X61689	B. burgdorferi ant
11	15.2	76.0	3107	C77447	Human ORFX ORF3002

12	15.2	76.0	7801	13	Q23500
13	15	75.0	1589	20	X84935
14	14.8	74.0	1321	20	X25655
15	14.8	74.0	1321	21	A59205
16	14.8	74.0	1422	21	C38943
17	14.8	74.0	1539	21	C42085
18	14.8	74.0	2372	20	X25664
19	14.8	74.0	2372	21	A59214
20	14.8	74.0	2422	20	X20844
21	14.8	74.0	2938	20	X25656
22	14.8	74.0	2938	21	A59206
23	14.8	74.0	3531	21	Z36787
24	14.8	74.0	6723	20	X13100
25	14.8	74.0	7582	20	X25665
26	14.8	74.0	7582	21	A59215
27	14.8	74.0	10636	20	X13011
28	14.8	74.0	138169	21	A34791
29	14.8	74.0	141589	21	F20913
30	14.8	74.0	141589	21	F21127
31	14.8	74.0	141589	21	F21152
32	14.8	74.0	141589	21	A35005
33	14.8	74.0	141589	21	A35030
34	14.8	74.0	146981	21	F21442
35	14.8	74.0	209273	21	F21457
36	14.6	73.0	818	13	Q32502
37	14.4	72.0	619	21	F11362
38	14.2	71.0	334	21	C34374
39	14.2	71.0	467	21	A30458
40	14.2	71.0	603	21	A10742
41	14.2	71.0	743	21	F13533
42	14.2	71.0	797	21	F08035
43	14.2	71.0	873	20	Z16753
44	14.2	71.0	882	19	V40176
45	14.2	71.0	1157	21	C34054

#### ALIGNMENTS

RESULT 1	
ID A50294	A50294 standard; DNA; 20 BP.
AC A50294:	
DT 20-NOV-2000	(first entry)
DE	Candida albicans CaESS1 gene-specific primer OM-221.
XX	
XX	CaESS1; ESS1; Infection; diagnosis; therapy; antifungal; fungicide;
KW	antiproliferative; cytostatic; antitumour; PCR primer; ss.
OS	Candida albicans.
XX	
PN	WO200050561-A2.
XX	
PD	31-AUG-2000.
XX	
PF	18-FEB-2000; 2000WO-US04203.
XX	
PR	23-FEB-1999; 99US-0121246.
XX	
PA	(HEAL-) HEALTH RES INC.
XX	
PI	Hanes SD, Devasahayam G, Chaturvedi V;
XX	
DR	WPI, 2000-565453/52.
XX	
PT	Novel Candida albicans gene, CaESS1 useful for identifying compounds
PT	that specifically bind to and/or inhibit CaESS1 and thus for treating
PT	Candida albicans infections and other life-threatening fungal
PT	infections -
XX	

Zymomonas xylose 1  
Human secreted pro  
Human endogenous r  
5' non coding and  
Arabidopsis thalia  
Arabidopsis thalia  
Human endogenous r  
Human endogenous r  
R-U5 region and pa  
Human secreted pro  
Human endogenous r  
Gag and partial po  
DNA sequence of th  
Enterococcus faeca  
Complete human end  
Human endogenous r  
Enterococcus faeca  
Human adenosine re  
Human ELAM-1 polyn  
Human ELAM-1 polyn  
Human low adenosin  
Human low adenosin  
Human adenosine re  
Human adenosine re  
Human ELAM-1 polyn  
Human factor-relat  
HCV NS4-NS5 clone  
Q32502  
F11362  
Aspergillus niger  
Arabidopsis thalia  
Human brain cDNA c  
Fragment #1 of C.  
Aspergillus oryzae  
Fusarium venenatum  
Human gene express  
Flavobacterium sp.  
Arabidopsis thalia

PS Claim 6; Page 38; 51pp; English.  
XX  
CC This is the nucleotide sequence of PCR primer OM-221, which is  
CC specific to the CaESS1 gene (see A50292), the Candida albicans  
CC homologue of Saccharomyces cerevisiae ESS1. OM-221 can be used  
CC with primer OM-216 (see A50293) for the specific amplification of  
CC CaESS1 DNA; they do not amplify ESS1 DNA or PIN1 (the human  
CC homologue of ESS1) DNA. Detection of CaESS1 nucleic acids can  
CC be used as a means for diagnosing C. albicans infection.  
XX  
SQ Sequence 20 BP; 6 A; 5 C; 6 G; 3 T; 0 other;  
  
Query Match 100.0%; Score 20; DB 21; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.076;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 caatgacgaggaacgtccg 20  
Db 1 caatgacgaggaacgtccg 20  
|||||  
  
RESULT 2  
A50292/c  
ID A50292 standard; DNA; 989 BP.  
XX  
AC A50292;  
XX  
XX 20-NOV-2000 (first entry)  
XX  
DE Candida albicans CaESS1 gene.  
XX  
XX CaESS1; ESS1; infection; diagnosis; therapy; antifungal; fungicide;  
XX KM antiproliferative; cytostatic; antitumour; ds.  
XX OS  
XX Candida albicans.  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS 256..789  
FT /\*tag= a  
XX  
XX MO200050561-A2.  
XX  
XX 31-AUG-2000.  
XX  
XX 18-FEB-2000; 2000WO-US04203.  
XX  
XX 23-FEB-1999; 9905-0121246.  
XX  
XX (HEAL-) HEALTH RES INC.  
XX  
XX Hanes SD, Devasahayam G, Chaturvedi V;  
XX  
XX WPI: 2000-565453/52.  
XX  
XX P-PSDB: Y95876.  
XX  
XX Novel Candida albicans gene, CaESS1 useful for identifying compounds  
XX PT that specifically bind to and/or inhibit CaESS1 and thus for treating  
XX PT Candida albicans infections and other life-threatening fungal  
XX PT infections -  
XX  
XX  
XX Claim 2; Fig 1A; 51pp; English.  
XX  
XX This is the complete nucleotide sequence of the CaESS1 gene from  
XX CC Candida albicans. The gene encodes a 177-amino acid protein,  
XX CC CaESS1 (see Y95876), which is the C. albicans homologue of  
XX CC Saccharomyces cerevisiae ESS1 (42% amino acid identity). The  
XX CC CaESS1 gene was isolated from a C. albicans genomic DNA library by  
XX CC functional complementation of a temperature-sensitive S. cerevisiae  
XX CC strain, ess1-194pts. CaESS1 nucleic acids, especially CaESS1-specific  
XX CC primers (see A50293-94) and probes to determine the presence of C.  
XX CC albicans in a sample or specimen. CaESS1 protein is a target  
XX CC for screening for antifungal and inhibitor compounds, useful for

CC treating or preventing C. albicans infections. CaESS1 DNA can also  
CC be used to generate diagnostic probes or primers for replicating or  
CC cloning C. albicans DNA. A CaESS1 inhibitor can be a compound  
CC which selectively inhibits growth of S. cerevisiae not containing an  
CC endogenous ESS1 gene but rather CaESS1 and uninduced PIN1 (a human  
CC homologue of ESS1) and/or preferably does not inhibit induced PIN1,  
CC e.g. does not inhibit S. cerevisiae not containing an endogenous  
CC ESS1 but rather induced PIN1. Compositions which inhibit PIN1  
CC are useful antiproliferatives e.g. antineoplastics, antitumour  
CC agents or anticancer agents.  
XX  
SQ Sequence 989 BP; 340 A; 174 C; 190 G; 285 T; 0 other;  
  
Query Match 100.0%; Score 20; DB 21; Length 989;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 caatgacgaggaacgtccg 20  
Db 970 caatgacgaggaacgtccg 951  
|||||  
  
RESULT 3  
T24451/c  
ID T24451 standard; cDNA to mRNA; 274 BP.  
XX  
AC T24451;  
XX  
XX 10-OCT-1996 (first entry)  
XX  
XX Human gene signature HUMGS06487.  
XX  
XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
XX KM human; cloning; mapping; non-biased library; diagnosis; detection;  
XX KW cell typing; abnormal cell function; ss.  
XX  
XX Homo sapiens.  
XX OS  
XX MO9514772-A1.  
XX  
XX 01-JUN-1995.  
XX  
XX 11-NOV-1994; 94WO-JP01916.  
XX  
XX 12-NOV-1993; 93JP-0355504.  
XX  
XX (MATS/) MATSUBARA K.  
XX PA (OKUB/) OKUBO K.  
XX  
XX PI Matsubara K, Okubo K;  
XX  
XX WPI: 1995-206931/27.  
XX  
XX  
XX Identifying gene signatures in 3'-directed human cDNA library - e.g.  
XX PT for diagnosis of abnormal cell function, by preparing cDNA that  
XX PT reflects relative abundance of corresp. mRNA in specific human  
XX PT tissues  
XX  
XX  
XX Claim 1; Page 1613; 2245pp; Japanese.  
XX  
XX A single-stranded DNA (or its complementary strand or the corresp.  
XX CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
XX CC given in T19001-T26837 and which is able to hybridise to part of  
XX CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
XX CC sequences were obtained from 3'-directed cDNA libraries prepared  
XX CC from various human tissues; synthesis of cDNA was initiated from the  
XX CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
XX CC untranslated sequence is unique to a particular mRNA species, almost  
XX CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
XX CC is constructed so as to reflect accurately the relative abundance of  
XX CC different mRNAs in the particular tissue from which it was derived.  
XX CC The appearance frequency of a given GS in a cDNA library can be

Fri Jun 8 09:21:27 2001

us-09-507-242-6.rng

determined (esp. using primers and probes derived from the GS sequences) as a means of diagnosing abnormal cell function or for recognizing different cell types.

Sequence 274 BP; 67 A; 71 C; 73 G; 49 T; 14 other;

Query Match 76.0%; Score 15.2; DB 16; Length 274;  
Best Local Similarity 85.0%; Pred. No. 42;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 caatgacgggaacgttcg 20  
|||||  
Db 226 CAATGACGGGAACATCTG 207

RESULT 4  
291619/C  
ID 291619 standard; DNA; 479 BP.

291619;

17-MAY-2000 (first entry)

E. coli promoter region flanking sequence, DPD2086 lower.

Promoter region; regulatory region identification; flanking sequence;  
cellular insulin; luciferase; thermostable lux gene complex;  
luxCDABE gene complex; crop protection chemical;

stress responsive regulatory region; ss.

Escherichia coli.

US6025131-A.

15-FEB-2000.

23-OCT-1996; 96US-0735545.

23-OCT-1996; 96US-0735545.

(DUPO) DU PONT DE NEMOURS & CO E I.

Larossa RA, Van Dyk TK;

WPI: 2000-181802/16.

Identification of regulatory regions modulated by a cellular insulin,  
such as crop protection chemicals -

Example: Column 29-30; 31pp; English.

This sequence is a flanking sequence for an E. coli promoter region.  
The invention relates to a method for identifying regulatory regions  
modulated by a cellular insulin, comprising: (a) creating a library of  
gene fusions of genomic DNA fused to a promoterless, luminescent reporter  
gene complex selected from a gene complex encoding and a luxCDABE gene  
complex in *Escherichia coli* to create fusion-containing strains in liquid media;  
(b) culturing individual gene fusion-containing strains in liquid media;  
(c) contacting the fusion-containing strains at a particular growth  
phase with a change in luminescence, the change in luminescence  
indicating that the fusion-containing strain includes a regulatory region  
modulated by the cellular insulin. The method is useful for identifying  
regulatory regions affected by cellular stress such as that created by  
crop protection chemicals. The method can detect promoters or stress  
responsive regulatory regions undetectable by prior art methods.

Sequence 479 BP; 132 A; 131 C; 104 G; 112 T; 0 other;

Query Match

76.0%; Score 15.2; DB 21; Length 479;

Best Local Similarity 85.0%; Pred. No. 45;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 caatgacgggaacgttcg 20  
|||||  
Db 218 CCACGACGAGAACGTTCCG 199

RESULT 5  
C04536  
ID C04536 standard; CDNA; 560 BP.

C04536;

06-OCT-2000 (first entry)

Human secreted protein 5' EST, SEQ ID NO: 8631.

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
gene therapy; chromosome mapping; ss.

Homo sapiens.

EP1033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-0200610.

26-FEB-1999; 99US-0122487.

(GEST) GENSET.

Dumas Milne Edwards J, Duclert A, Giordano J;

WPI: 2000-500381/45.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
diagnostic, forensic, gene therapy and chromosome mapping procedures -  
Claim 1: SEQ ID 8631; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from  
mRNAs encoding secreted proteins. No ORF has yet been conclusively  
identified within the present sequence. The 5' ESTs were prepared from  
total human RNAs or polyA+ RNAs derived from 30 different tissues (UTR)  
sequences usually correspond mainly to the 3' untranslated region (UTR)  
of the mRNA because they are often obtained from oligo-dT primed cDNA  
libraries. Such ESTs are not well suited for isolating cDNA sequences  
derived from the 5' ends of mRNAs and even in those cases where longer  
cDNA sequences have been obtained, the full 5' UTR is rarely included.  
5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
They are used to obtain upstream regulatory sequences and to design  
expression and secretion vectors.

Sequence 560 BP; 113 A; 185 C; 172 G; 90 T; 0 other;

Query Match 76.0%; Score 15.2; DB 21; Length 560;  
Best Local Similarity 85.0%; Pred. No. 45;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 caatgacgggaacgttcg 20  
|||||  
Db 274 caatgacgggaacgttcg 293

RESULT 6  
C42788/C  
ID C42788 standard; DNA; 603 BP.

```
XX AC C42788;
XX 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 36852.
XX XX
XX KM Hybridisation assay; genetic mapping; gene expression control;
XX KM protein identification; signal transduction pathway;
XX KM metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PE 25-FEB-2000; 2000EP-0301439.
XX XX
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 28-APR-1999; 99US-0130891.
XX PR 30-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 04-MAY-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132464.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 07-MAY-1999; 99US-0132487.
XX PR 11-MAY-1999; 99US-0132863.
XX PR 14-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 18-MAY-1999; 99US-0134370.
XX PR 19-MAY-1999; 99US-0134768.
XX PR 20-MAY-1999; 99US-0134941.
XX PR 21-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
XX PR 10-JUN-1999; 99US-0138847.
XX PR 14-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139492.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139455.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.
XX PR 18-JUN-1999; 99US-0139459.
XX PR 18-JUN-1999; 99US-0139460.
XX PR 18-JUN-1999; 99US-0139461.
XX PR 18-JUN-1999; 99US-0139462.
XX PR 18-JUN-1999; 99US-0139463.
XX PR 18-JUN-1999; 99US-0139753.
XX PR 21-JUN-1999; 99US-0139817.
XX PR 22-JUN-1999; 99US-0139899.
XX PR 23-JUN-1999; 99US-0140353.
XX PR 23-JUN-1999; 99US-0140354.
XX PR 24-JUN-1999; 99US-0140695.
XX PR 28-JUN-1999; 99US-0140823.
XX PR 29-JUN-1999; 99US-0140991.
XX PR 30-JUN-1999; 99US-0141287.
XX PR 01-JUL-1999; 99US-0141842.
XX PR 02-JUL-1999; 99US-0142154.
XX PR 06-JUL-1999; 99US-0142055.
XX PR 08-JUL-1999; 99US-0142390.
XX PR 09-JUL-1999; 99US-0142803.
XX PR 12-JUL-1999; 99US-0142920.
XX PR 13-JUL-1999; 99US-0142977.
XX PR 14-JUL-1999; 99US-0143542.
XX PR 15-JUL-1999; 99US-0143624.
XX PR 16-JUL-1999; 99US-0144005.
XX PR 16-JUL-1999; 99US-0144085.
XX PR 16-JUL-1999; 99US-0144086.
XX PR 19-JUL-1999; 99US-0144325.
XX PR 19-JUL-1999; 99US-0144331.
XX PR 19-JUL-1999; 99US-0144332.
XX PR 19-JUL-1999; 99US-0144333.
XX PR 19-JUL-1999; 99US-0144334.
XX PR 19-JUL-1999; 99US-0144335.
XX PR 20-JUL-1999; 99US-0144352.
XX PR 20-JUL-1999; 99US-0144632.
XX PR 20-JUL-1999; 99US-0144884.
XX PR 21-JUL-1999; 99US-0144814.
XX PR 21-JUL-1999; 99US-0145086.
XX PR 21-JUL-1999; 99US-0145088.
XX PR 22-JUL-1999; 99US-0145085.
XX PR 22-JUL-1999; 99US-0145087.
XX PR 22-JUL-1999; 99US-0145089.
XX PR 22-JUL-1999; 99US-0145192.
XX PR 23-JUL-1999; 99US-0145145.
XX PR 23-JUL-1999; 99US-0145218.
XX PR 23-JUL-1999; 99US-0145224.
XX PR 26-JUL-1999; 99US-0145276.
XX PR 27-JUL-1999; 99US-0145913.
XX PR 27-JUL-1999; 99US-0145918.
XX PR 27-JUL-1999; 99US-0145919.
XX PR 28-JUL-1999; 99US-0145951.
XX PR 02-AUG-1999; 99US-0146386.
XX PR 02-AUG-1999; 99US-0146388.
XX PR 02-AUG-1999; 99US-0146389.
XX PR 03-AUG-1999; 99US-0147038.
XX PR 04-AUG-1999; 99US-0147204.
XX PR 04-AUG-1999; 99US-0147302.
XX PR 05-AUG-1999; 99US-0147192.
XX PR 05-AUG-1999; 99US-0147260.
XX PR 06-AUG-1999; 99US-0147303.
XX PR 06-AUG-1999; 99US-0147416.
XX PR 09-AUG-1999; 99US-0147493.
XX PR 09-AUG-1999; 99US-0147935.
XX PR 10-AUG-1999; 99US-0148171.
XX PR 11-AUG-1999; 99US-0148319.
XX PR 12-AUG-1999; 99US-0148341.
XX PR 13-AUG-1999; 99US-0148565.
XX PR 13-AUG-1999; 99US-0148684.
XX PR 16-AUG-1999; 99US-0149368.
XX PR 17-AUG-1999; 99US-0149175.
XX PR 18-AUG-1999; 99US-0149426.
XX PR 20-AUG-1999; 99US-0149722.
XX PR 20-AUG-1999; 99US-0149723.
XX PR 20-AUG-1999; 99US-0149929.
XX PR 23-AUG-1999; 99US-0149902.
XX PR 23-AUG-1999; 99US-0149930.
```

PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 76.0%; Score 15.2; DB 21; Length 603;  
Best Local Similarity 85.0%; Pred. No. 46;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 caatgacgggaaccttcg 20  
Db 274 CAATGACGAGAAACGTGCTG 255

RESULT 7  
ID C48389 standard; DNA; 1062 BP.  
XX C48389;  
XX 18-OCT-2000 (first entry).  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 57303.  
DE

KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
XX EPI033405-A2.  
PD 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.

PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 11-OCT-1999; 99US-0158369.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 22-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 76.0%; Score 15.2; DB 21; Length 1062;  
Best Local Similarity 85.0%; Pred. No. 49;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 caatgacggaacgtccg 20  
|||||  
Db 733 CAATGACGAGAACGTCTG 714

RESULT 8  
F13745/c  
ID F13745 standard; CDNA; 1315 BP.

XX F13745;  
AC  
XX  
XX 13-MAR-2001 (first entry)  
XX  
XX  
DE Aspergillus oryzae EST SEQ ID NO:6268.

KW Multiple gene expression; filamentous fungal cell; EST;  
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
KW culture condition; environmental stress; spore morphogenesis;  
KW metabolic pathway engineering; catabolic pathway engineering; ss.  
XX  
OS Aspergillus oryzae.



```

XX XX WO200056762-A2.
XX PN
XX XX
XX XX 28-SEP-2000.
XX PD
XX XX
XX XX 22-MAR-2000; 2000WO-US07781.
XX PF
XX XX 22-MAR-1999; 99US-0273623.
XX PR
XX XX (NOVO ) NOVO NORDISK BIOTECH INC.
XX PA (NOVO ) NOVO NORDISK AS.
XX PA
XX XX Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX PI WPI: 2000-594572/56.
XX XX
XX XX Monitoring differential expression of genes in filamentous fungal cells
XX PT uses fluorescence-labeled nucleic acids isolated from the cells and a
XX PR substrate of expressed sequence tags -
XX PS
XX XX Claim 88: Page 2577-2578; 3161pp; English.
XX XX
XX XX The present invention describes a method for monitoring differential
XX CC expression of genes in a first filamentous fungal (FF) cell relative to
XX CC expression of the same genes in one or more second filamentous fungal
XX CC cells. The method uses fluorescence-labeled nucleic acids isolated from
XX CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX CC are used in the methods for monitoring differential expression of genes
XX CC in a first filamentous fungal (FF) cell relative to expression of the
XX CC same genes in one or more second filamentous fungal cells. Monitoring
XX CC the global expression of genes from FF cells allows the production
XX CC potential of the microorganisms to be improved. New genes may be
XX CC discovered, possible functions of unknown open reading frames can be
XX CC identified and gene copy number variation and stability can be
XX CC monitored. The expression of genes can be used to study how FF cells
XX CC adapt to changes in culture conditions, environmental stress, spore
XX CC morphogenesis, recombination, metabolic or catabolic pathway engineering.
XX CC Using ESTs provides several advantages over genomic or random cDNA
XX CC clones including elimination of redundancy as one spot on an array
XX CC equals one gene or open reading frame, and organisation of the
XX CC microarrays based on function of the gene products to facilitate
XX CC analysis of the results. F07478 to F1247 represents ESTs from Fusarium
XX CC venenatum; F11248 to F11853 represents ESTs from Aspergillus niger;
XX CC F11854 to F14878 represents ESTs from Aspergillus oryzae; and F14879 to
XX CC F15337 represents ESTs from Trichoderma reesei, which are all
XX CC specifically claimed in the present invention.
XX CC
XX XX Sequence 1315 BP; 305 A; 370 C; 262 G; 378 T; 0 other;
XX SO

```

Query Match 76.0%; Score 15.2; DB 21; Length 1315;  
 Best Local Similarity 85.0%; Pred. No. 50;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

OY 1 caatgacggaagcgttcgcg 20
   ||| ||| ||| ||| ||| |||
Db 954 CAAGACACGGAACGCTCCG 935

```

RESULT 9  
 X61690  
 ID X61690 standard; DNA: 1473 BP.  
 XX  
 XX AC X61690;  
 XX  
 XX DT 19-JUL-1999 (first entry)  
 XX  
 XX DE B. burgdorferi antigenic protein coding sequence, t752.nt.  
 XX  
 XX KW Antigenic protein; vaccine; Lyme disease; Infection; detection; ss.  
 XX  
 XX PA Borrelia burgdorferi.  
 XX

```

PN PM WO9859071-A1.
XX XX
XX PD 30-DEC-1998.
XX XX
XX XX 18-JUN-1998; 98WO-US12718.
XX PF
XX XX 03-SEP-1997; 97US-0057483.
XX PR 20-JUN-1997; 97US-0050359.
XX PR 22-JUL-1997; 97US-0053344.
XX PR 22-JUL-1997; 97US-0053377.
XX XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (MEDI-) MEDIMUNE INC.
XX XX
XX PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX XX
XX XX WPI: 1999-189980/16.
XX DR P-PSDB: Y19993.
XX XX
XX XX New isolated Borrelia burgdorferi nucleic acids - used to develop
XX PT products for the diagnosis, prevention and treatment of diseases
XX PR caused by Borrelia, particularly Lyme disease
XX PS
XX XX Claim 1: Page 154; 275pp; English.
XX XX
XX XX This sequence encodes a Borrelia burgdorferi (Bb) protein of the
XX CC invention, which is suitable for use in a vaccine. The Bb polypeptides
XX CC can be used in vaccines for eliciting protective antibodies to members of
XX CC the Borrelia genus, particularly for the use against Lyme disease in
XX CC humans and animals. They can be used for preventing or attenuating an
XX CC infection caused by a member of the Borrelia genus. The products can also
XX CC be used for detection of members of the Borrelia genus.
XX XX
XX SO Sequence 1473 BP; 468 A; 166 C; 290 G; 549 T; 0 other;

```

Query Match 76.0%; Score 15.2; DB 20; Length 1473;  
 Best Local Similarity 85.0%; Pred. No. 51;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

OY 1 caatgacggaagcgttcgcg 20
   ||| ||| ||| ||| ||| |||
Db 300 caatgacggaagcgttcgcg 319

```

RESULT 10  
 X61689  
 ID X61689 standard; DNA: 1527 BP.  
 XX  
 XX AC X61689;  
 XX  
 XX DT 19-JUL-1999 (first entry)  
 XX  
 XX DE B. burgdorferi antigenic protein coding sequence, f752.nt.  
 XX  
 XX KW Antigenic protein; vaccine; Lyme disease; Infection; detection; ss.  
 XX  
 XX PA Borrelia burgdorferi.  
 XX  
 XX OS WO9859071-A1.  
 XX PN  
 XX PD 30-DEC-1998.  
 XX  
 XX PF 18-JUN-1998; 98WO-US12718.  
 XX  
 XX PR 03-SEP-1997; 97US-0057483.  
 XX PR 20-JUN-1997; 97US-0050359.  
 XX PR 22-JUL-1997; 97US-0053344.  
 XX PR 22-JUL-1997; 97US-0053377.  
 XX  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX PA (MEDI-) MEDIMUNE INC.  
 XX

PI Choi GH, Erwin AL, Hanson MS, Iathigra R;  
 XX MPI: 1999-189980/16.  
 DR P-PSDB; Y19992.  
 XX  
 PT New isolated *Borrelia burgdorferi* nucleic acids - used to develop  
 PT products for the diagnosis, prevention and treatment of diseases  
 PT caused by *Borrelia*, particularly Lyme disease  
 PS  
 PS Claim 1: Page 153-154; 275pp: English.  
 XX  
 CC This sequence encodes a *Borrelia burgdorferi* (Bb) protein of the  
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
 CC can be used in vaccines for eliciting protective antibodies to members of  
 CC the *Borrelia* genus, particularly for the use against Lyme disease in  
 CC humans and animals. They can be used for preventing or attenuating an  
 CC infection caused by a member of the *Borrelia* genus. The products can also  
 CC be used for detection of members of the *Borrelia* genus.  
 CC  
 SQ Sequence 1527 BP; 496 A; 168 C; 296 G; 567 T; 0 other;  
 Query Match 76.0%; Score 15.2; DB 20; Length 1527;  
 Best Local Similarity 85.0%; Pred. No. 51;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 OY 1 caatgacggaagaacttcgcg 20  
 II |||||IIIIII |||||  
 Db 354 cactgacggagaaagtctcg 373  
 RESULT 11  
 C77447/c  
 ID C77447 standard; cDNA; 3107 BP.  
 XX  
 AC C77447;  
 XX  
 DT 08-FEB-2001 (first entry)  
 DE  
 DE Human ORFX ORF3002 polynucleotide sequence SEQ ID NO:6003.  
 XX  
 KW Human: open reading frame: ORFX; detection: cytostatic; hepatotropic;  
 KW vulnervary; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200058473-A2.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 31-MAR-2000; 2000MO-US08621.  
 XX  
 PR 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shinkets RA, Leach M;

DR MPI: 2000-602362/57.  
 DR P-PSDB; B43238.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 PS  
 PS Claim 5: Page 5182-5184; 5507pp: English.  
 XX  
 CC C74446 to C77606 encode the proteins given in B40237 to B43397, which  
 CC represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnervary;  
 CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective; osteopathic;  
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;  
 CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The  
 CC sequences can be used for determining the presence of or predisposition  
 CC to, or preventing or treating pathological conditions associated with an  
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
 CC used to treat cancers, proliferative disorders, neurodegenerative  
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.  
 CC  
 SQ Sequence 3107 BP; 546 A; 965 C; 981 G; 613 T; 2 other;  
 Query Match 76.0%; Score 15.2; DB 21; Length 3107;  
 Best Local Similarity 85.0%; Pred. No. 55;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 OY 1 caatgacggaagaacttcgcg 20  
 II |||||IIIIII |||||  
 Db 2459 CACTGACGAGGAACCTTCGC 2440

RESULT 12.  
 ID Q23500 standard; DNA; 7801 BP.  
 XX  
 AC Q23500;  
 XX  
 DT 26-AUG-1992 (first entry)  
 DE  
 DE Zymomonas xylose isomerase expression plasmid pX1PX.  
 XX  
 KW xylA; xylB; pZM1A; pyruvate decarboxylase; promoter; marker gene;  
 KW chloramphenicol resistance; ss.  
 KW  
 XX Zymomonas mobilis.  
 OS  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FH CDS 1618..2940  
 FT /tag- a  
 FT /product= xylose\_isomerase  
 FT /note= "see R22720"  
 FT 3013..4468  
 FT /tag- b  
 FT /standard\_name= xylB  
 FT /note= "xylokinase"  
 FT 1521..1592  
 FT /tag- c  
 FT /label= pyruvate\_decarboxylase\_promoter  
 FT 1531..1543  
 FT /tag- d  
 FT 1559..1568  
 FT /tag- e  
 FT -10\_signal  
 FT -35\_signal  
 FT -10\_signal

FT	misc_feature	1..219	
FT		/*tag= f	
FT	misc_feature	/phenotype= chloramphenicol_resistance	
FT		7361..7801	
FT		/*tag= g	
FT		/phenotype= chloramphenicol_resistance	
FT	misc_feature	5414..7096	
FT		/*tag= h	
FT		/label= pM21A	
FT		/note= "when included in a plasmid this region	
FT		allows replication in <i>zymomonas</i> "	
XX			
PN	JF04066090-A.		
XX			
PD	02-MAR-1992.		
XX			
PF	05-JUL-1990;	90JP-0176420.	
XX			
PR	05-JUL-1990;	90JP-0176420.	
XX			
PA	(ENER-) SHIN ENERGY SANGYO.		
XX	(NENR-) NENRYOYO ALCOHOL KAIHAT.		
XX			
DR	WPI: 1992-120685/15.		
XX	P-PSDB: R22720.		
PT	Vector for expression of heterogenes in <i>zymomonas</i> sp.		
PT	used for expression of xylose isomerase for prodn. of ethanol from		
PT	xylose and/or fructose		
XX			
PS	Disclosure: Fig 1; 14pp: Japanese.		
XX			
CC	This is the sequence of plasmid pX1PY. The plasmid can be used to		
CC	transform <i>zymomonas</i> cells. The transformants can be cultured to		
CC	express the xylA/B genes. Alcohol production from such		
CC	transformants is comparable to that from yeast.		
XX			
CC			
SO	Sequence 7801 BP: 1959 A; 1910 C; 1918 G; 2014 T; 0 other;		
XX			
Query Match	76.0%; Score 15.2; DB 13; Length 7801;		
Best Local Similarity	85.0%; Pred. No. 62;		
Matches 17; Conservative	0; Mismatches 3; Indels 0; Gaps 0		
QY	1 caatagcggaacgttcgcg 20		
DB	4410 ccacgacgagaacgttcgcg 4429		
RESULT 13			
X84935/c			
ID	X84935 standard; DNA: 1589 BP.		
XX			
AC	X84935:		
XX			
DT	30-JUL-1999 (first entry)		
XX			
DE	Human secreted protein gene No. 3.		
XX			
KW	Human; secreted protein; fusion protein; gene therapy; protein therapy;		
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;		
KW	developmental abnormality; foetal deficiency; blood; allergy; renal; ds;		
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;		
KW	inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;		
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;		
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;		
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.		
XX			
OS	Homo sapiens.		
XX			
PN	W09924836-A1.		
XX			
XX			
PD	20-MAY-1999.		

[illegible]

```

OS Human endogenous retrovirus.
XX
PN WO9902696-A1.
XX
PD 21-JAN-1999.
XX
PF 06-JUL-1998; 98WO-FR01442.
XX
PR 07-JUL-1997; 97FR-0008815.
XX
PA (INMR ) BIO MERIEUX.
XX
PI Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;
XX
DR WPI: 1999-120897/10.
XX
PT New nucleic acid sequences from human endogenous retrovirus-W -
XX expressed exclusively in placenta and useful in diagnosis and
XX therapy of autoimmune disease, and abnormal or failed pregnancy
XX
PS Claim 1; Page 48-49; 106pp; French.
XX
CC This sequence represents clone cl.6A2 of the human endogenous retrovirus
CC (HEV) W genome. The nucleic acids, their fragments or peptides encoded
CC by them are markers of autoimmune disease (e.g. multiple sclerosis,
CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
CC dependent diabetes and related pathologies) and of abnormal or
CC unsuccessful pregnancy and can be used as chromosomal markers for
CC susceptibility to these conditions, or proximity markers of genes
CC associated with this susceptibility.
XX
SQ Sequence 1321 BP; 365 A; 352 C; 277 G; 327 T; 0 other:
XX
XX
Query Match 74.0%; Score 14.8; DB 20; Length 1321;
Best Local Similarity 88.9%; Pred. No. 83;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 2 aatgacgggaaacgttc 19
1 ||| ||||| |||||
Db 602 agtgatgggaaacgttc 619
RESULT 15
A59205 A59205 standard; DNA; 1321 BP.
AC A59205;
XX
DT 07-NOV-2000 (first entry)
XX
DE 5' non coding and partial gag fragment of HERV-W from human genome.
XX
KW Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
XX gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
OS Homo sapiens.
XX
PN WO200043521-A2.
XX
PD 27-JUL-2000.
XX
PF 21-JAN-2000; 2000WO-FR00144.
XX
PR 21-JAN-1999; 99FR-0000888.
XX
PA (INMR ) BIO MERIEUX.
XX
PI Paranhos-Bacala G, Mallet F, Voisset C;
XX
DR WPI: 2000-499229/44.
XX
PT New nucleic acid from human endogenous retrovirus, useful e.g. for

```

PT diagnosis of autoimmune diseases and complications of pregnancy,  
PT contains at least part of the gag gene -  
XX  
PS  
XX Disclosure: Page 42-43; 53pp; French.  
XX  
CC The present sequence represents an endogenous retroviral nucleic acid  
CC fragment, which is associated with an autoimmune disease, and is  
CC integrated into the human genome. The fragment is originally derived  
CC from a novel retrovirus, human endogenous retrovirus W (HERV-W). The  
CC HERV-W retrovirus is associated with autoimmune disease, failure of  
CC pregnancy or disorders of pregnancy. The nucleic acid fragment, or  
CC proteins derived from it, are useful for diagnosis of autoimmune  
CC disease (specifically multiple sclerosis) and for monitoring pregnancy  
CC The nucleic acid fragments may also be used for in situ labelling of  
CC isolated chromosomes, while the transcription product can be used to  
CC study or monitor T cell proliferation in vitro.  
XX  
XX Sequence 1321 BP; 365 A; 352 C; 277 G; 327 T; 0 other;

	Query Match	Score 14.8;	DB 21;	Length 1321;
	Best Local Similarity	88.9%	Pred. No. 83;	
Matches	16;	Conservative	0;	Mismatches
			2;	Indels
			0;	Gaps
Qy	2	aatgacggaacgcttc	19	
Db	602	agtgatgggaacgcttc	619	

Search completed: June 8, 2001, 06:21:27  
Job time: 6151 sec

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

This Page Blank (usp...)